University of London

FUNCTIONAL IMAGING STUDIES OF VISUAL-AUDITORY INTEGRATION IN MAN

Richard Kirk Lewis

Ph.D. Thesis

University College London

2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This thesis investigates the central nervous system's ability to integrate visual and auditory information from the sensory environment into unified conscious perception. It develops the possibility that the principle of functional specialisation may be applicable in the multisensory domain. The first aim was to establish the neuroanatomical location at which visual and auditory stimuli are integrated in sensory perception. The second was to investigate the neural correlates of visual-auditory synchronicity, which would be expected to play a vital role in establishing which visual and auditory stimuli should be perceptually integrated.

Four functional Magnetic Resonance Imaging studies identified brain areas specialised for: the integration of dynamic visual and auditory cues derived from the same everyday environmental events (Experiment 1), discriminating relative synchronicity between dynamic, cyclic, abstract visual and auditory stimuli (Experiment 2 & 3) and the aesthetic evaluation of visually and acoustically perceived art (Experiment 4).

Experiment 1 provided evidence to suggest that the posterior temporo-parietal junction may be an important site of crossmodal integration. Experiment 2 revealed for the first time significant activation of the right anterior frontal operculum (aFO) when visual and auditory stimuli cycled asynchronously. Experiment 3 confirmed and developed this observation as the right aFO was activated only during crossmodal (visual-auditory), but not intramodal (visual-visual, auditory-auditory) asynchrony. Experiment 3 also demonstrated activation of the amygdala bilaterally during crossmodal synchrony. Experiment 4 revealed the neural correlates of supramodal, contemplative, aesthetic evaluation within the medial fronto-polar cortex. Activity at this locus varied parametrically according to the degree of subjective aesthetic beauty, for both visual art and musical extracts.

The most robust finding of this thesis is that activity in the right aFO increases when concurrently perceived visual and auditory sensory stimuli deviate from crossmodal synchrony, which may veto the crossmodal integration of unrelated stimuli into unified conscious perception.

Acknowledgements

My three years here at the Zeki lab have been the most demanding, challenging and exciting of my life so far. I would like to express my deepest gratitude to Professor Zeki for taking me on in the first place, for having faith in my experimental paradigms, for nurturing my skills as a scientist, an intellectual and a gentleman and for maintaining an environment of good humour throughout. I genuinely feel I have developed in leaps and bounds in every manner and will be set loose on the world as a young scientist armed with all the necessary practical skills, self-belief and knowledge to make a success of whatever I put my heart and soul into.

I would not have made it on my own without the friendship and advice of the various mentors and colleagues that Professor Zeki has astutely selected in recent times. Each and every one that I have had the pleasure to work with has been instrumental in my development, but of these I will reserve special praise for the most influential. To Andreas 'the love doctor' Bartels, who was (and still is) a Zeki post-doc of extraordinary ability, I thank for his help in designing and executing my first experiment. To Dr Matthew Self, my guru of all things Matlab and robust experimental paradigms, I owe possibly the greatest debt of gratitude, as he single-handedly taught me all the programming, SPM and study design skills that enabled me to make such a success of my time here. To Oliver Hulme and Barrie Roulston, whom along with Dr Self form the SOAS crew with whom I enjoyed many a pleasant social, I credit my intact sanity. Their friendship and advice through the ups and downs of this research degree led me safely away from the maw of self-doubt.

I would like to acknowledge the enthusiastic support of my friends, who constantly reminded me of how grateful I should be to be doing something fascinating with my life. I am immensely grateful to my girlfriend Rebecca Jolley, for her support, patience and understanding, and for accommodating my workaholism.

Finally, I wish to thank Phil and Virginia Lewis, who have provided the sturdy platform from which all their children have launched into academic careers in which they find themselves enjoying considerable success. The innumerable sacrifices, which enabled the provision of a stable, loving home, the best education anyone could wish for and the freedom to explore and experience the world, have not gone unnoticed and will never be forgotten. My eternal gratitude goes to them for all they have provided and for making the best interests of their children their own best interests.

Table of Contents

List of Figures

INTRODUCTION

METHODS and RESULTS

Experiment 1

Experiment 2

List of Tables

List of Publications

Publication 1: R.K. Lewis, S. Zeki. Visual-auditory asynchrony detection in the right frontal operculum. Program No. 528.11. 2004 Abstract Viewer/Itinerary Planner. Washington, DC: Society for Neuroscience, 2004. Online.

Publication 2: R.K. Lewis, S. Zeki. Crossmodal, but not intramodal, asynchrony activates the right anterior frontal operculum.2005 Abstract Viewer/Itinerary Planner. Washington, DC: Society for Neuroscience, 2005.....abstract accepted

Publication 3: R.K. Lewis, S. Zeki. Bilateral amygdala activation during inter-sensory temporal covariation ..... in prep.

Part 1:

INTRODUCTION

1.1 Multisensory perception

Despite significant progress in understanding of how sights and sounds are processed under unimodal conditions (i.e. individually) within visual and auditory regions of the mammalian brain, relatively little is known about how visual and auditory information is integrated into unified perception. This mechanism is referred to throughout this thesis as *crossmodal integration*. Furthermore, it is not even known where in the brain regulation of crossmodal integration takes place, a process described in this thesis as *multisensory grouping*. Such a brain area must be capable of distinguishing between visual and auditory stimuli that should undergo crossmodal integration (i.e. those that arise from the same environmental event), from those that should remain separated in perception (i.e. those that arise from different environmental events).

It seems reasonable to assume that the purpose of sensory perception is to gain knowledge about the world (Zeki and Bartels, 1999). In order for knowledge to be gained from sensory information, perception must involve more than just the reception of data, what is received must be recognised and organised (Kant, 1787) - it is an active not a passive process. It is therefore assumed that both the crossmodal integration of visual-auditory stimuli emanating from the sensory environment and the mechanism regulating multisensory grouping are active processes and should therefore result in increased activation at the relevant functionally specialised brain area.

1.1.1 Visual-auditory bimodal perception

All mammals are equipped with multiple sensory channels through which the environment can be experienced. Unimodal perception describes conscious awareness through just one of these sensory modalities, whilst bimodal perception describes conscious awareness through two sensory modalities concurrently. Visual and auditory unimodal sensory perception imparts several advantages over the other sensory modalities. Vision, audition (and olfaction) share the advantage over the other senses of enabling the perception of distal phenomena, and in dispensing with the need for proximity, permit sensory experience at a safe distance. Mammalian evolution has witnessed a declining reliance upon olfaction in favour of vision and audition (Jerison, 1970), to which considerable cortical resources have become devoted, particularly in the primate brain (Poremba et al., 2003). Consequently, primate visual and auditory senses

are able to extract considerably more reliable and detailed information from the environment than chemoreception. Arguably the most important of these advances is the ability to rapidly process complex, dynamic information from which the behaviour of environmental agents and events can be evaluated, patterns established and predictions about future behaviours made.

During the last century, the vast majority of neurobiological research into perception has concentrated almost exclusively on the anatomy and physiology of unimodal sensory processing. This has revealed some fascinating insights into how visual and, to a lesser extent, auditory parts of the sensory cortex have become specialised to extract information from the sights and sounds that emanate from the environment. Conversely, how sensory information from the different senses might be combined to produce the fully integrated and seemingly unified state of conscious awareness that we as humans take for granted has been largely neglected.

For static objects, there is a natural correspondence between visual and tactile perception of edges and surfaces, a compatibility that lends itself to effortless unified bimodal experience when a seen object is explored by hand. Such tight correspondence in visual-auditory perception is only made possible through dynamic stimulation (Jones, 1981). One of the most important characteristics of bimodal perception is that both unique and redundant information is provided from which sensory concepts can be derived. Examples of unique unimodal sensory phenomena that have no equivalent in the other sensory system's repertoire are colour in the visual system and pitch in the auditory system. On the other hand, certain agents and/or events in the sensory environment can be characterised both by qualities discerned from analysis of the reflected light *and* of the emitted acoustic signature.

1.1.2 Intermodal invariances in bimodal perception

Certain fundamental equivalences between visual and auditory cues arising from the same environmental object enable them to be used entirely interchangeably. This is reflected in the behaviour of human infants as young as 3 months old (Piaget, 1953). In his *Critique of Pure Reason*, Immanuel Kant proposed that it is a condition of having any sensory experience at all that the world is experienced as including relations of cause and effect, being ordered in time and that perceived objects are related spatially to each other (Kant, 1787). In his discussion of the "transcendental aesthetic", Kant points

out that all sensory "intuitions" must necessarily occur in the context of time and space. Several universally valid observations arise from this discussion, which the brain can use to reliably indicate whether visual and auditory attributes originate from a single environmental entity, or from separate sources:

> "Different times are not simultaneous but successive"
>
> "Different spaces are not successive but simultaneous"
>
> (Kant, 1787)

Brain areas involved in multisensory grouping can therefore use coincidence in time and space between the visual and auditory 'effects' arising from an environmental 'cause' to govern crossmodal integration. Brain areas involved in this process must be able to distinguish between visual and auditory percepts that arise at the same time and from the same part of space, from those that do not. When such equivalences are present between visual and auditory cues, it is appropriate to consider them as a single entity. When these equivalences are absent, they should be perceived as separate and unrelated. These are the basic principles of multisensory grouping. When an occurrence in the sensory environment is perceived as a single entity, equivalences in time and space form the hinge upon which complementary sensory information can be hung, resulting in the formation of percepts with a greater depth of detail than would be possible based on the visual or auditory parts alone.

In addition to time and space, other *intermodal invariances*, i.e. features that are analogous irrespective of the sensory modality through which they are conveyed, include intensity, rate and rhythmic structure (Lewkowicz, 2000). Intermodal invariances comprise the set of equivalences on the basis of which dynamic changes in a visual stimulus can be perceptually integrated with changes in an auditory stimulus.

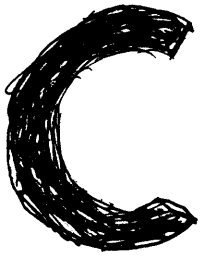
Rhythm is an excellent example of an intermodal invariance that enables visual and auditory stimuli to be perceptually united in conscious awareness and has been defined as:

> "the perception of temporal form or pattern in which individual members repeated periodically are consistently varied in any one or more of their qualitative or quantitative attributes"
>
> (Ruckmick, 1927)

Rhythm is extremely important in bimodal visual-auditory perception because not only do matching sequences of stimulus changes across the senses provide a vital indication of intermodal invariance for multisensory grouping, but as perception of order is also conveyed, it becomes possible to anticipate what is to follow (Fraisse, 1981). Furthermore, in addition to providing predictive power over the future course of sensory events within a single modality, during synchronous rhythmic stimulation in the visual and auditory domains, the visual stimulus enables the future outcome of the auditory stimulus to be anticipated, and *vice versa*, leading to a mutual reinforcement of the intermodal invariance and the combined predictive power.

1.1.3 A naturally-occurring, rhythmic, audiovisual phenomenon

The sight and sound of ocean waves crashing up the beach is an example of a natural, rhythmic, audiovisual phenomenon. Both qualitatively distinct and correlated sensory information are captured and conveyed through the sensory apparatus in such a way as to produce a merged, unified, bimodal percept. As suggested above, bimodal perception is much richer than the sum of its component parts, providing a fuller sensory description than can be obtained from just the sight or sound of the wave in isolation. In addition to the visual and auditory impressions, the crossmodal equivalences that can be implicitly derived from the dynamic features shared by the visual and auditory cues provide for a deeper understanding of the nature of a wave and result in the formation of a more complex wave concept. The merged percept of a wave can be dissected into visual and auditory, but also 'crossmodal' components that equate sensory changes between the covarying visual and auditory stimuli (figure 1).



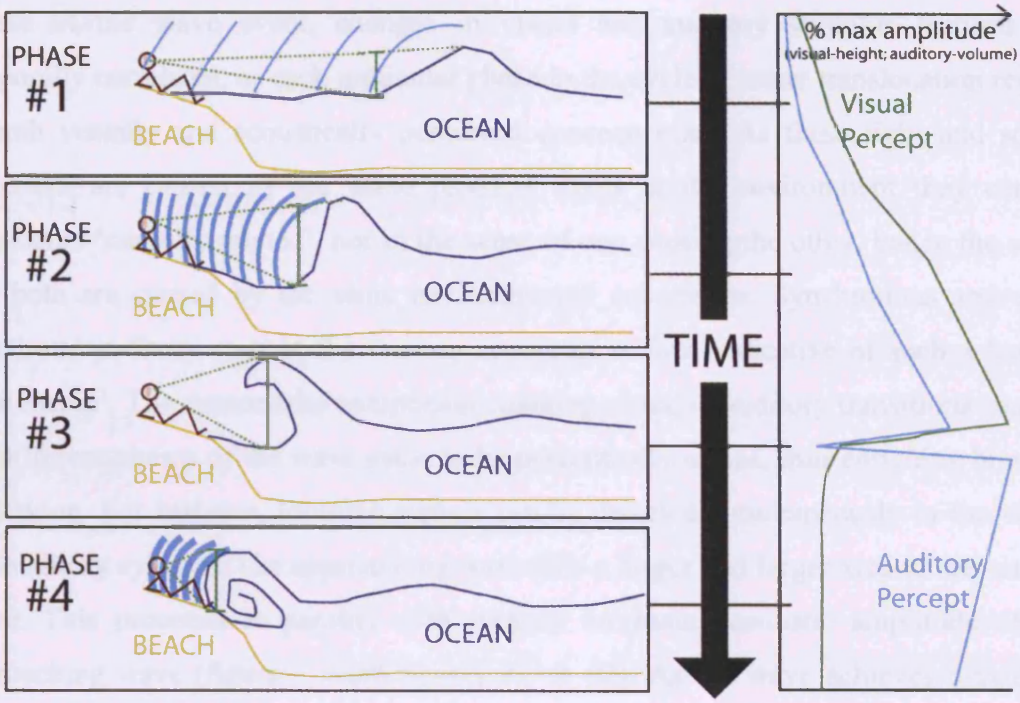


Figure 1: Example of a naturally-occurring, rhythmic, audiovisual phenomenon. Environmental stimuli that produce both visual and auditory sensory perturbations can provide information that is more than the sum of its parts. For instance, from the perspective of an observer sitting on the shoreline, an ocean wave can be described by the visual transitions, the auditory transitions and additional crossmodal information can be gleaned from the correlation between coincident occurrences across visual and auditory transitions. This is described in more detail below.

As figure 1 illustrates, the visual sequence of events consists of the approaching wave becoming gradually taller and taller as it nears the beach and at the last moment it rears up, curls over and crashes down onto the beach. The visual percept thus provides dynamic information regarding changes in its height (the dynamic visual feature focused on in figure 1), but also information regarding its colour, shape and uniformity – which could not be gleaned from merely listening to the same stimulus. The auditory stimulus arising from the same phenomenon comprises the gradual increase in pitch and amplitude (the dynamic acoustic feature focused on in figure 1) of the sounds created by the translocation of water as the wave nears the shore. When the wave reaches full height there is a momentary silence preceding the clamour of the wave's collapse onto the beach, followed by the crackling, effervescence of the resulting wash. The auditory percept provides its own unique information regarding the mass, force and speed of water movement that could not necessarily be visually discerned, even providing information regarding events disguised beneath the water's surface. Throughout the

course of the wave event, changes in visual and auditory stimulus features are temporally correlated, as each particular phase in the cycle of water translocation results in both visually and acoustically perceived consequences. As these light and sound emissions are caused by the same physical event in the environment they can be considered 'causally related', not in the sense of one causing the other, but in the sense that both are caused by the same environmental occurrence. Synchronous arrival of visual and auditory cues at the sensory apparatus is thus indicative of such a 'causal relationship'. The crossmodal component equating visual to auditory transitions enables the different phases of the wave cycle to be perceptually united, thus enriching bimodal perception. For instance, looming signals can be detected simultaneously in the visual and auditory systems. The approaching wave fills a larger and larger area of retinotopic space. This proceeds in parallel with steadily increasing acoustic amplitude of the approaching wave (figure 1, from phases #1 → #2). As the wave achieves maximum height, just before it breaks, the increase in encroachment on new areas of retinotopic space momentarily ceases, which is coincident with a brief pause in the acoustic emission (phase #3). Finally, the ensuing auditory clamour is synchronised with (NB caused by), the visually perceived collapse of the high, transparent, blue-green wall into effervescent white foam (phase #4). Unified bimodal perception thus becomes possible when such intermodal invariances can be established between a pair of visual and auditory stimuli.

1.2 Thesis aims

In essence, this thesis has been built around the concept that unified bimodal perception may be supported by specialised brain areas dedicated to integrating and regulating the integration of visual and acoustic stimuli in the environment. The studies described in this thesis were undertaken to investigate where in the brain: 1) crossmodal perceptual integration of, 2) multisensory grouping between and 3) supramodal evaluation of visual and auditory stimuli is achieved. As we shall see in the historical survey, the psychophysical evidence demonstrating the existence and utility of basic visual-auditory integration is strong (section 2.2), but the brain areas involved in actually enabling such integration are currently very poorly understood.

1.2.1 The problem of reconciling unified sensory perception with functional specialisation

Visual functional specialisation, which will be reviewed briefly in the historical survey (section 2.1), has posed a problem in neuroscience, namely the method by which separately processed components of a single object can be processed apart, yet perceived together. To take the specific example of the various essential nodes that divide the visual brain into its constituent submodalities: if the form, colour and motion of a single object are processed in separate visual areas, how can these visual components be brought back together in perception? An essential node can be described as a specific neuroanatomical site within parallel sensory processing pathways, at which, beyond a certain threshold, conscious perception of the attribute for which that node is specialised occurs (Zeki and Bartels, 1999).

This problem can easily be extended from unimodal to unified bimodal perception. How can the visual and auditory components pertaining to a single object, distributed so broadly across the carefully segregated sensory systems, be brought together in perception? Given that all brain regions have reciprocal connections with other areas, and that there is no single region at which all neuronal circuits ultimately terminate, this 'binding' question is not a simple one. Many theories have been proposed to explain how separately processed features can be associated with a single object, including both theories of neuronal resonance (Singer, 1999) and dissonance (Bartels and Zeki, 1998). None has elucidated an all-encompassing theory to explain

precisely how the fruits of these disparate elements of sensory processing are brought together in our seamless and seemingly unified perceptual experience.

This brings us to the central question raised by this debate. In the face of compelling evidence indicating a modular organisation in the visual brain, with separated brain areas specialised for the extraction of different features or sub-modalities, might there also be specialised regions of the brain responsible for extracting intermodally invariant information from visual and auditory sensory streams? The pathways and essential nodes for various aspects of visual and (to a lesser extent) auditory processing have been well characterised, permitting us to examine the similarities and differences between anatomical areas supporting unimodal and bimodal perception. Broadly speaking there are two ways in which crossmodal integration might be achieved. On the one hand this may be induced by reciprocal modulation of 'unimodal' sensory territories. On the other, visual and auditory information might be brought together at a third neuroanatomical location, dedicated to crossmodal integration.

If the coarsely defined lessons from neurology are taken as a starting point: depending on the size and neuroanatomical location of the lesion, the effects of damaging cortical regions responsible for the processing and perception of visual (occipital) and auditory (temporal) stimuli are partial or total blindness and deafness, respectively. Such observations converge with findings in anatomical and electrophysiological animal experimentation and, in more recent times, with non-invasive human imaging techniques, to culminate in the irrefragable certainty that the extraction of visual and auditory information is dependent on structures within the occipital lobes and superior temporal structures, respectively. This begs the simple question, addressed from several different angles in this study, namely, how are the outputs of visual and auditory stimulus processing, which are performed in broadly separated parts of the brain, recombined to form a single unified percept? In other words, the competing hypotheses are as follows: does integrated conscious awareness result from the merging of these two physically separated sensory streams through a) convergence at dedicated crossmodal brain regions, or b) through some form of heightened intercommunication between unimodal brain regions? (Figure 2).

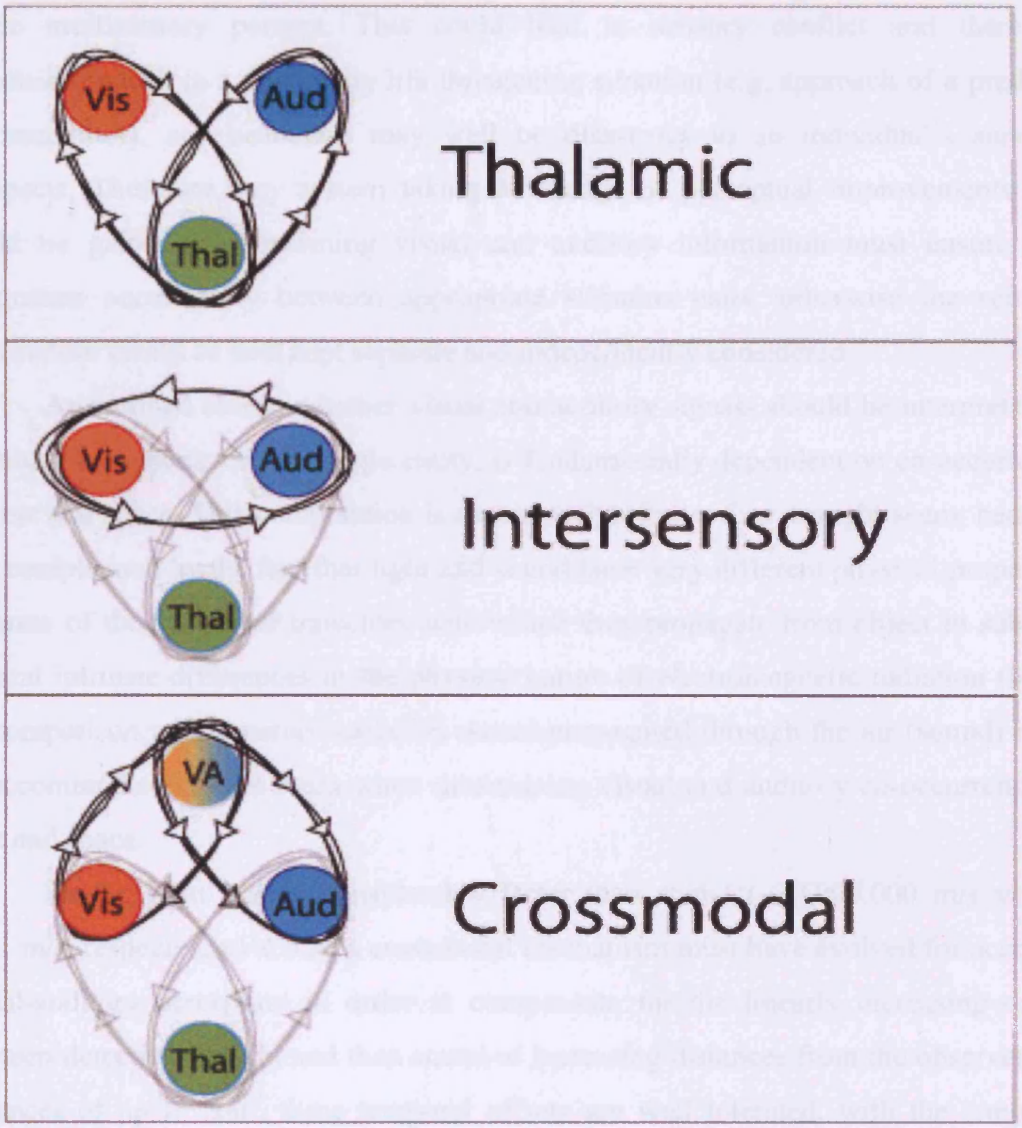


Figure 2: The possible neuronal connectivity underlying visual and auditory sensory integration in conscious perception. This could be achieved by a thalamic mechanism, direct communication between unimodal sensory cortices, or indirect interaction at committed crossmodal areas to result in unified bimodal perception. (Vis – visual cortex; Aud – auditory cortex; VA – multisensory cortex; Thal – thalamus)

1.2.2 The best strategy for foolproof visual-auditory integration

A vitally important question that must be asked of a system that brings visual and auditory information together into perceptual unity is how it can distinguish between unrelated multisensory events that co-occur by chance, from those that co-occur because they actually result from the same physical disturbance in the environment. The worst possible scenario is for visual and auditory perturbations relating to completely different environmental entities to be erroneously merged into a

single multisensory percept. This could lead to sensory conflict and therefore indecision, which in a potentially life threatening situation (e.g. approach of a predator or competitor), any hesitation may well be disastrous to an individual's survival prospects. Therefore, any system taking advantage of perceptual improvements that might be gained by combining visual and auditory information must ensure that integration occurs only between appropriate stimulus pairs, otherwise the sensory information would be best kept separate and independently considered.

As outlined above, whether visual and auditory signals should be interpreted as separate phenomena, or as a single entity, is fundamentally dependent on co-occurrence in time and space. This computation is not as straightforward as it might seem, because it is complicated by the fact that light and sound have very different physical properties in terms of the speed and trajectory with which they propagate from object to subject. Several intrinsic differences in the physical nature of electromagnetic radiation (light) by comparison with pressure variation waves propagated through the air (sound) must be accommodated in the brain when determining visual and auditory co-occurrence in time and space.

Firstly, light travels considerably faster than sound (~30,000,000 m/s versus ~331 m/s, respectively) and so a crossmodal mechanism must have evolved for accurate visual-auditory perception in order to compensate for the linearly increasing offset between detection of light and then sound at increasing distances from the observer. At distances of up to 20m, these temporal offsets are well tolerated, with the temporal discrepancy between the arrival of visual and auditory stimuli discounted, enabling them to be subjectively perceived as arriving simultaneously (reviewed in section 2.2.2). As the inability of young children to grasp that lightening and thunder are not actually separate meteorological phenomena, but temporally delayed sensory markers of exactly the same distant event testifies - the offset between detection of light and sound at greater distances from the original event are not well tolerated.

Secondly, light travels in straight, unbroken lines and the visual sensory apparatus is organised to preserve and take advantage of the high degree of spatial resolution that can be derived from analysis of the arrangement of reflected light falling on the retina. This retinotopic organisation is critical to the processing of the visual scene, providing precise information regarding the arrangement of object features and the relative positioning of objects in space. Conversely, sound waves do not travel in straight lines, but are able to pass through, around and reflect off intervening obstacles.

Whilst having the advantage of providing acoustic sensory information emanating from visually occluded and thus otherwise unperceivable phenomena, the spatial information available from acoustic signals is often comparatively poor as a result. It is in the spectrotemporal domain that the most valuable acoustic information can be extracted and this is reflected in the tonotopic organisation of the auditory system. Tonotopy in the auditory system describes the decomposition of the detected sounds into constituent frequency bands at the cochlea, which is preserved at the level of the primary auditory cortex and beyond, allowing efficient extraction of the rich spectrotemporal information.

In a busy environment, such as the woods, a jungle or the high street, many temporal coincidences between unrelated visual and auditory stimuli originating from approximately the same region of space would be expected to occur merely by chance. With the accuracy of implicit judgements about whether it is appropriate for the sensory apparatus to consider visual and auditory stimuli as single or separate phenomena so easily compromised in this way, a more reliable hallmark for visual-auditory integration must be determined.

This may indicate the need for a mechanism capable of reliably distinguishing between chance coincidence and 'meaningful' coincidence, flexible enough to operate at various observer-target distances and tolerant of spatial discrepancies that might arise in cases where the sound reaches the ear as an echo. In order to confidently indicate that visual and auditory stimuli arise from a common origin, this critical cue must be influenced neither by such slight temporal offsets, nor spatial discrepancies between the two cues. The only logical solution to this problem is that *consistent temporal covariation,* between dynamic visual and auditory cues, is the most efficient and reliable indication of common cause and therefore suitability for crossmodal integration. This principle is the foundation stone upon which experiments 2 and 3 were built. In other words consistent synchrony, or in the words of Horace Barlow (Barlow, 1986) "suspicious coincidences", between visual and auditory temporal variations should permit their combination in conscious perception. Any violation of consistent synchrony between visual and auditory stimulus changes should by the same token result in the prevention of integrated bimodal perception, to ensure that they remain segregated unimodal percepts.

1.2.3 Crossmodal integration past and present

There is a popular, but controversial view, originally proposed on the basis of visual deficits in animal lesion studies (Ungerleider and Mishkin, 1982), but recently suggested to apply equally in the auditory domain (for review: Hall, 2003), that sensory processing can be split into two parallel pathways: the ventral stream - specialised to determine *what* an object is, and the dorsal stream - responsible for *where* it is in extrapersonal space. To date, there has been a distinct bias in crossmodal integration studies favouring the latter topic, i.e. neural mechanisms supporting performance enhancements in visual-auditory object localisation (reviewed in: Stein and Meredith, 1993). As for crossmodal integration of stimulus features during combined audiovisual stimulation, the question has almost invariably been asked in the context of improved speech detection (Callan et al., 2001; Calvert et al., 1999; Calvert et al., 2000; Olson et al., 2002; Sams et al., 1991). This thesis is concerned neither with crossmodal improvements in cue localisation, nor speech detection. In all experiments speech stimuli were avoided in order to investigate the phylogenetic precursor to any brain area that might have evolved to specifically support crossmodal integration or synchronicity detection to aid speech detection. Therefore non-linguistic visual and auditory stimuli were used in all experiments and the position of stimuli was never varied. Functional Magnetic Resonance Imaging (fMRI) was used to scan subjects during exposure to complex, dynamic stimuli of various descriptions to understand where in the brain:

1) stimuli might be crossmodally integrated during bimodal as compared to unimodal visual and auditory conditions,
2) bimodal synchronicity might be monitored for accurate multisensory grouping, and
3) supramodal stimulus evaluation might take place.

Part 2:

HISTORICAL SURVEY

2.1 Cerebral functional specialisation

Over two hundred years of neurological and neurobiological analysis have converged in the last few decades to indicate that the sensory brain is composed, in part at least if not wholly, of many discrete neuroanatomical units; each specialised to perform a specific function. Upon first considering the possibility that the sensory brain is compartmentalised into anatomical sub-divisions, each specialised for the processing and perception of colour, pitch, motion, timbre, objects, melodies, faces and so forth, this may seem to conflict with the daily experience of such sensory features as a fully integrated whole. Counterintuitive as it may be, this is in fact the now broadly accepted scheme of organisation in the visual cortex and may well prove to also be the case in the organisation of the auditory cortex.

Considerable resistance and even ridicule was meted out to exponents of the principles of this theory throughout the course of its development. The often fervent opposition to such a concept may well have stemmed from the damning rejection of the science of 'phrenology' in the early nineteenth century. Just after the turn of the century the eminent anatomist, Dr John Gordon (1786-1818), speaking at the Surgeons' Square in Edinburgh publicly dismissed phrenology as, 'thorough quackery from beginning to end'. The godfathers of phrenology, Gall and Spurtzheim, had proposed that the cerebral topography of the brain might be anatomically divided into functionally discrete 'intellectual organs' (Gall and Spurtzheim, 1809). In principle this was not far from the truth, despite the rather arbitrary nature of the actual functions they ascribed to each area. The true 'quackery' stemmed from their belief that indications of such brain specialisations could be gleaned from the bumps and undulations of the overlying skull. Thus the proverbial baby, in the form of the first proposition of a system of functional localisation in the brain, was thrown out with the bathwater, i.e. the ludicrous supposition that palpation of the cranium could reveal the extent to which such functions are developed in an individual.

With phrenology officially exorcised from creditable scientific circles, so too was the principle of functional localisation, spurring a hasty return to the prevailing view at the time, that brain function was subserved by mass action to which all areas of the uniform and homogenous 'syncitium' contributed equally (Flourens, 1842). Ever since, and often in the face of convincing neurological support, any suggestion of functional localisation was rejected as nothing more than phrenology. Indeed, the

concept of mass action has survived through to modern models of brain function, which remained popular throughout much of the twentieth century (Lashley, 1946).

The modern day foundations of functional specialisation were built on rigorous animal electrophysiology, which bolstered support for the observations, oft reported but invariably overlooked, from neurological investigations of brain-damaged human subjects. The former delineated regional variation in neuronal response selectivity for certain categories of sensory stimuli over others. The latter demonstrated a neurological division of labour by correlating the site of a cortical lesion with the profile of faculties that are preserved and eliminated, in order to define the critical function that might be sub-served by the damaged tissue. Conclusions reached through convergence between animal and patient studies have since been corroborated by comparative anatomical connectivity studies and non-invasive human imaging techniques.

2.1.1 Key discoveries

The first evidence for functional localisation in the human brain was derived from the astute observations made by Pierre Paul Broca and divulged at a meeting of the Société d'Anthopologie, having just completed the post-mortem of a patient known as Tan (Broca, 1861). When Tan was first admitted to hospital he had lost the ability to speak, he later became paralysed down the right side of his body, his vision began to fail and his intelligence was also affected towards the end of his days (Zeki, 1993). Broca's autopsy revealed an extensive softening of the whole left frontal lobe and beyond, but with a clearly discernable 'primitive seat', where the most advanced and extensive lesion tissue was located, in the middle part of the left frontal lobe, just superior to the Sylvian fissure. Broca drew a link between the anatomical location of the earliest lesion and the original functional deficit that Tan presented with when he was first admitted to hospital, namely poverty of speech. The notion that Tan's aphasia might have been induced by damage to a discrete region of left frontal cortex was greeted with considerable scepticism by the experts of the time (Gratiolet, 1861). It is perhaps not surprising that the functional localisation was resisted initially because, at that time, the notion that the loss of a specific function could result from damage to a certain discrete part of the brain was the exception and not the rule.

Some years later, evidence to support cortical localisation for another function, the production of body movements, surfaced from the systematic study of dog motor

cortex (Fritsch and Hitzig, 1870). It was found that by electrically stimulating the anterior part of the frontal convexity 'combined muscular contractions of the opposite half of the body' could be obtained. Most importantly a direct relationship between cerebral architecture and function was demonstrated, which encouraged neurologists to think of distinct cortical areas as having specific functions.

The strongest drive behind the final acceptance of functional localisation was provided by research into visual perception. That the occipital lobe was critical for visual perception was first implicated by lesion work in the dog (Munk, 1881). Munk demonstrated that 'mind-blindness' followed certain cortical lesions, whereby dogs could see, but could not recognise objects in the visual world. Around the same time neurological examination of brain damaged patients led to the description of two types of visual agnosia: apperceptive agnosia, where the deficit was described as a failure to form complex visual percepts, and associative agnosia, characterized by the inability to attach appropriate meaning to visual objects (Lissauer, 1890). This led to the notion of a two-stage process in vision whereby 'primary' visual cortex was thought to passively receive the 'visual impressions' from the retina, but conscious perception was not possible without the second stage, which involved the formation of connections between the content of the perceptions with other conceptions, in the surrounding 'association' cortex. Soon after, in a study of centrally blind patients, Henschen went even further to propose a point-to-point relationship between adjacent regions of damaged primary visual cortex located in the occipital lobe and the areas of retinotopic space affected by blindness (Henschen, 1894). The functional division between primary and associative sensory cortices was supported by studies of cortical myelination during early development (Flechsig, 1901) and of the cytoarchitecture of the mature brain (Brodman, 1905), which both independently described parcellation of the cortical mantle into sub-divisions delineating a border between primary and association cortex.

2.1.2 Visual functional specialisation

Early neurological observations suggesting that lesions damaging specific cortical areas could lead to an inability to perceive (a) colour, a condition known as achromatopsia (MacKay, 1888; Verrey, 1888; Zeki, 1990) and (b) motion, or akinetopsia (Poezl and Redlich, 1911; Riddoch, 1917; Zeki, 1991; Zihl et al., 1983) raised the possibility that discrete regions of extrastriate cortex might process specific attributes of the visual scene. These findings struggled to gain acceptance as they opposed the broadly accepted doctrine of the time, that there was only one visual area, Henschen's 'cortical retina', located bilaterally along the calcarine sulci (Holmes, 1945; Zeki, 1993).

The concept of functional specialisation, first implied by this clinical evidence that mysteriously vanished from the literature, was not broadly accepted until pioneering electrophysiological investigations in rhesus macaque monkey provided incontrovertible evidence that discrete regions of extrastriate brain areas were activated specifically by colour (Zeki, 1973) and motion (Dubner and Zeki, 1971; Zeki, 1974). The region responsible for colour processing, located on the ventral occipital surface, became known as V4 and for motion processing as V5, located on the posterior bank of the superior temporal sulcus. Both areas receive direct and highly convergent input from the striate cortex (V4: Zeki, 1973; V5: Cragg, 1969; Zeki, 1969; Zeki, 1971). These findings ultimately led to the composition of a theory of functional specialisation in the visual cortex (Zeki, 1978), which revolutionised our understanding of the modularity of brain function.

The ultimate confirmation of functional specialisation in man was not possible until the advent of non-invasive imaging techniques, such as Positron Emission Tomography (PET) and later functional Magnetic Resonance Imaging (fMRI). These techniques would enable demonstration of functional human brain activity at anatomical locations homologous to the extrastriate visual areas determined physiologically in the monkey. Using such techniques the human colour centre was identified at the fusiform gyrus (Lueck et al., 1989), which has become known as the V4 complex based on more recent investigation (Bartels and Zeki, 2000). This was in accordance with the neuropathological data regarding the locus of brain lesions that results in cerebral achromatopsia (Zeki, 1990). Shortly afterwards, functional activation of the human motion centre, V5, was demonstrated on the ventro-lateral convexity at the junction

between the occipital and temporal lobes (Watson et al., 1993), once again confirming the lesion anatomy in akinetopsia (Zeki, 1991; Zihl et al., 1983; Zihl et al., 1991). Having provided converging evidence for visual functional specialisation from monkey electrophysiology, human neurology, and finally in the living, perceiving human brain using non-invasive neuroimaging, the principle is now surely impossible to refute.

2.1.3 Auditory functional specialisation

In the late nineteenth century the superior temporal cortex was identified as the neuroanatomical location for auditory sensory processing based on animal studies (Ferrier, 1876). Early analyses of human cytoarchitecture established that the transverse temporal (Heschl's) gyrus was the site of human primary auditory cortex (hPAC) (Brodman, 1909; von Economo, 1930). More recent studies of superior temporal cytoarchitecture have established that the hPAC resides in the medial two thirds of Heschl's gyrus (Galaburda and Sanides, 1980; Rademacher et al., 1993) at which early-evoked potentials to a variety of sounds are invoked 8-10ms after sound onset (Celesia, 1976; Liegeois-Chauvel et al., 1994; Liegeois-Chauvel et al., 1991). Scientific progress in auditory research has lagged considerably behind visual research and so, consequently, nothing absolutely definitive can be said regarding the functional specialisations of the auditory cortex beyond the hPAC.

It was not until the late twentieth century that subdivisions and connectivity of extra-primary auditory areas began to emerge in the owl monkey and the macaque (Fitzpatrick and Imig, 1980; Merzenich and Brugge, 1973; Morel et al., 1993; Morel and Kaas, 1992; Rauschecker et al., 1997). These studies revealed that the core fields (primary auditory cortex) could be distinguished from the adjacent rostrolateral field (R or RL) and were surrounded in a belt-like arrangement by other distinct auditory representations. Just as in the visual cortex, boundaries between adjacent auditory regions are marked by mirror reversals in neuronal tuning preference, but rather than reversals in retinotopic space, the reversals are in frequency preference, thus resulting in several separate tonotopic maps of auditory frequency space (Morel et al., 1993; Rauschecker et al., 1997). Furthermore, the core area demonstrates the most organised and sharply tuned tonotopy, whilst progression through successive belt areas yields response functions of increasing complexity and broadening tuning functions. This can be likened to the increasing size of visual receptive fields during the progression from

primary to extrastriate visual areas (Smith et al., 2001). The sensitivity to less specific frequency bandwidths may be indicative of functional specialisation in the auditory system, as sensitivity to less specific regions of space (i.e. increased receptive field size) is necessary for V4 and V5 to perform colour and motion analysis, respectively (Essen and Zeki, 1978; Komatsu et al., 1992). Furthermore, in addition to the more coarsely defined tonotopy, lateral belt areas are also responsive to complex acoustic features such as the direction and extent of frequency modulations (Rauschecker et al., 1995; Whitfield and Evans, 1965). In fact, the more distant belt areas do not respond to simple sounds at all, reacting only to complex sounds (Manley and Mueller-Preuss, 1978). Taken together, these observations seem to imply a degree of auditory functional specialisation, but without providing definitive proof for the involvement of specific areas responsible for the processing and perception of individual, submodal, auditory features.

The planum temporale lies directly behind Heschl's gyrus and is considered, based on human cytoarchitectural evidence, to be the secondary auditory area (Galaburda and Sanides, 1980). Recent, auditory neuroimaging experiments have established that the planum temporale is significantly more active in response to frequency and amplitude modulated sounds than to pure tones or noise (Giraud et al., 2000; Hart et al., 2003) and yet more active when the carrier signal is a harmonic-complex tone rather than a single tone (Hall, 2002).

Perhaps the best neurological evidence for functional specialisation for high-level, complex, sound sequence processing comes from a condition known as amusia. It describes an acquired loss of the ability to perceive music, whilst the ability to perceive other environmental sounds is maintained (Griffiths et al., 1997). In keeping with neurological observations suggesting that amusia results from damage to the right cerebral hemisphere, functional imaging studies have implicated the right superior temporal cortex in the perceptual analysis of melodies as compared to auditory stimuli of similar acoustic characteristics (e.g. Zatorre et al., 1994).

Other studies have indicated that changes in rhythm lead to activations in Broca's area and adjacent parts of the insula in the left hemisphere, pointing to a role for these structures in the sequencing of auditory input (Patel et al., 1998). Behavioural studies of brain-damaged patients indicate that left superior temporal areas may be involved in the processing of local melodic intervals in musical sequences, whereas

right frontotemporal circuits are involved in the perception of global pitch contour (Liegeois-Chauvel et al., 1998).

Acquired auditory agnosia is often but not invariably associated with bilateral lesions of the superior temporal lobe (Griffiths, 1999). It seems that even with intact superior temporal auditory cortices, patients with frontal lobe lesions can demonstrate deficits in processing and perceiving patterned sound, suggesting an important role for frontal structures in these processes (Griffiths et al., 2000).

Although this appraisal of the auditory literature is not exhaustive, it is clear that the evidence supporting the notion of functional specialisation in the auditory system is far from complete, with only scattered and disjointed support for various theories regarding the relative contributions of different auditory areas to the various aspects of spectrotemporal sound analysis. Indeed, precisely what functions can be ascribed to the various subdivisions of auditory cortex remains a hotly contested issue of debate (e.g. Hall, 2003).

2.1.4 Evidence for crossmodal functional specialisation

The overarching question addressed in this thesis is, if visual and auditory sensory processing is performed within broadly separated neuroanatomical locations, how can the appropriate visual and auditory features be reassembled to produce the holistic visual-auditory perception of everyday experience? In the rest of this historical review, evidence that addresses this issue from several scientific disciplines has been collated. First studies from the psychophysical literature are considered which describe behavioural performance advantages that can be achieved through visual-auditory integration in man (section 2.2). The main factors affecting these processes and a selection of crossmodal phenomena that can result from conflicting visual and auditory information are outlined. Then cortical territories are identified where responses to both visual and auditory stimuli have been detected in electrophysiological investigations, which might therefore be involved in the generation of behaviours that could reasonably be assumed to rely upon crossmodal integration (section 2.3). The neurological literature is then briefly examined for indications of the common sites of human conditions in which normal visual-auditory perception is in some way disrupted (section 2.4). This is followed by a brief review of the most relevant functional imaging papers

addressing key issues pertinent to the pursuit of neural correlates of crossmodal integration and synchronicity (section 2.5).

There is a rather unfortunate habit in the discourse of visual-auditory crossmodal integration research, whereby the evidence for interactions between any and all of the senses are cited together without differentiating between which pairs of senses are integrated. In other words references to interactions between all combinations of visual, auditory, somatosensory, vestibular and proprioceptive stimuli are often cited together as evidence to support the concept of crossmodal integration irrespective of the senses involved. Effort has been made to avoid perpetuating this trend to give a more accurate representation of the evidence that is of specific relevance to visual-auditory interactions and integration.

2.2 The psychophysics of visual-auditory perception

Concurrent sensory processing of visual and auditory stimuli in man was first approached by cognitive psychologists in the hope of better understanding the nature of attention and the consequences of dividing it (reviewed: Miller, 1991). The capacity of one sensory system to modify, that is to enhance, degrade or change, perception in another system is known as intersensory bias in the psychological literature (e.g. Bernstein et al., 1969; Bernstein et al., 1970; Hershenson, 1962; O'Leary and Rhodes, 1984; Sekuler et al., 1997). A classic example of perceptual enhancement conveyed through intersensory bias or, to use the more specific term visual-auditory integration, is the cocktail party effect (Sumby and Pollack, 1954). This phenomenon embodies many of the essential conditions under which information derived from these two senses can be combined to produce a sensory signal of superior fidelity to either in isolation. It demonstrates a commonly experienced intersensory bias that takes advantage of the synergy between the visual and auditory speech signals. In a room with significant background noise it can be very difficult to hear what a person is saying if their mouth is obscured. This is because the speech information provided by the sight of a speaker's lips is combined with the information derived from auditory analysis of the speaker's voice, thus improving signal to noise. In fact, the perceptual advantage that this conveys has been calculated, indicating that the signal is enhanced by an amount equivalent to increasing the auditory signal by 15-20 dB SPL (Sumby and Pollack, 1954). This is just one of several examples of visual-auditory facilitation that are outlined below.

2.2.1 The redundant target effect

When subjects monitor information presented in two different modalities and are required to respond to targets presented in either modality, they are usually more accurate with simultaneously rather than separately presented targets (e.g. Colquhoun, 1975). A popular model of response accuracy, based on signal detection theory (Green and Swets, 1979), explained this finding in terms of independent processing of the different channels, followed by summation of the resulting sensory activations, which was in turn compared with a response criterion. This was later elaborated upon to allow differential target weighting to contribute to the overall activation (Kinchla, 1977; Kinchla and Collyer, 1974). In addition to improved accuracy, responses to such combined, or bimodal targets, are also faster than to individually presented, or

unimodal, targets (e.g. Miller, 1982; Miller, 1986). As either of the targets could have elicited the response there is redundancy in the stimuli, and so the reaction time improvement in these tasks became known as the redundant-target effect (RTE).

Early models put forward to explain this phenomenon, such as the energy summation theory (Bernstein et al., 1970) and the preparation enhancement model (Nickerson, 1973), could not fully account for the RTE in all circumstances. Shortcomings in such models were particularly apparent in experiments in which visual and acoustic stimuli were always presented, but some pairs were targets and others were distracters (Miller, 1982). In other words they were not flexible enough to accommodate the task 'context'. The *race model* had more success in such scenarios, by assuming independent processing of both sensory streams and proposing that whichever reaches the output stage first triggers the response. *Statistical facilitation* (Raab, 1962) provides the foundations for this purely mathematical theory, under the assumption that processing times for the two channels are randomly distributed. It simply implies that there is a higher probability that the reaction time for one or other component of the bimodal stimulus falls into the lower half of the reaction time distribution than for the unimodal stimulus. However, the observed reaction times in bimodal RTE experiments *do* violate the race model assumptions, leading to support for the competing coactivation hypothesis (Miller, 1982; Miller, 1986; Miller, 1991; Ulrich and Giray, 1986). As opposed to the race model, the *coactivation model* suggests that the processing pathways of both modalities converge at some point where processing efficiency is increased by bimodal input. This increase in efficiency accounts for both the improved accuracy and faster responses to bimodal than unimodal stimuli. The only remaining difficulty arises in finding a consensus of opinion as to whether this coactivation occurs at a sensory (Miniussi et al., 1998; Schroger and Widmann, 1998), a decisional (Miller, 1991), or a motor (Ulrich and Giray, 1986) level (see figure 3).

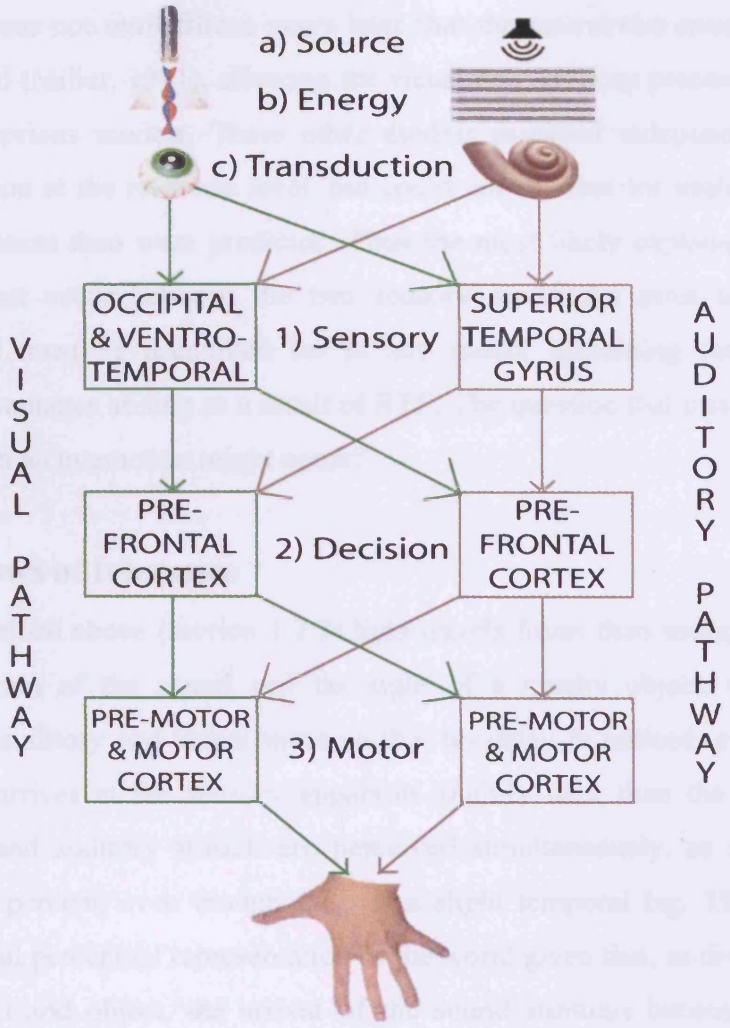


Figure 3: Models explaining the redundant target effect. This involves visual and auditory sources (a), releasing stimulus energies simultaneously (b), which are transduced into neuronal signals at the retina and cochlea, respectively (c). Concurrent processing of these bimodal stimuli results in faster and more accurate responses than when either of the same stimuli is presented in isolation. The coactivation model provides the most compelling explanation for the redundant target effect, but proponents differ in opinion as to whether visual-auditory interactions occur at the (1) sensory (Miniussi et al., 1998; Schroger and Widmann, 1998), (2) decisional (Miller, 1991), or (3) motor (Ulrich and Giray, 1986) level.

The early reluctance to consider the possibility of interaction between visual and auditory systems in the pursuit of an explanation for crossmodal behavioural performance advantages is curious. This may have been partly to do with prevailing belief that the anatomical separation of visual and auditory cortices precluded any interaction at the perceptual level. It may have simply related to the logical approach of creating forward models, where most basic explanations are considered first, only searching for more complex explanations elsewhere once these have been exhausted.

Either way, it was not until fifteen years later that the *interactive coactivation* model was put forward (Miller, 1991), allowing for visual and auditory processing to interact sooner than previous models. These other models assumed independent processing before summation at the response level, but could not account for trials that produced more fast responses than were predicted. Thus the most likely explanation is that the interactions must occur between the two sensory modalities prior to the response production and must be accounted for in any model explaining these crossmodal behavioural advantages arising as a result of RTE. The question that this raises is where in the brain such an interaction might occur?

2.2.2 Windows of tolerance

As described above (section 1.2.2) light travels faster than sound. Therefore for unified perception of the sound and the sight of a nearby object, the brain must coordinate the auditory and visual input so that no delay is noticed, even though the sound always arrives at the sensory apparatus slightly later than the light. In other words, visual and auditory stimuli are perceived simultaneously, as a single bound visual-auditory percept, even though there is a slight temporal lag. This is vital to a coherent bimodal perceptual representation of the world given that, as distance increases between subject and object, the arrival of the sound stimulus becomes increasingly delayed with respect to that of the light stimulus by 3ms for every 1m (Sugita and Suzuki, 2003). The Sugita and Suzuki study demonstrated that crossmodal temporal integration of this nature is tolerant of stimulus onset asynchronies (SOAs) of between 50-100ms, suggesting that visual-auditory SOAs will go unnoticed at distances up to 20m. It has been suggested that this is achieved by using information about distance that is supplied by the visual system to calibrate simultaneity (Sugita and Suzuki, 2003; but see: Lewald and Guski, 2004).

For a visual and an auditory event to be registered as a single coherent percept, some degree of spatial and temporal coincidence is vital. For the reasons outlined above, the requirement for this is not absolute, rather they must fall within a flexible, dynamic 'window' of coincidence defined along spatiotemporal dimensions (Wallace et al., 2004). A certain amount of spatial discrepancy can be tolerated so long as the temporal offset between visual and auditory events is consistent (Radeau, 1994).

Equally, so long as other features are shared by the bimodal stimuli (such as spatial origin and duration), a degree of temporal asynchrony can also be accommodated.

Classic psychophysical studies have revealed that there are subtle differences in the exact temporal window of tolerance whether subjects are required to detect the point of subjective simultaneity (PSS), or whether they are required to detect just noticeable differences (JND) in the onset times of a visual and an auditory stimulus (Spence et al., 2003). One study used stimuli consisting of a ball bouncing on a surface accompanied by an impact sound, and found that they were perceived to occur synchronously so long as the sound did not precede the visible bounce by more than 80ms and did not follow the visible bounce by more than 140ms (Summerfield, 1979). A more recent study using static, discrete crossmodal stimulus pairs found that they were consistently perceived as simultaneous within a slightly different range from auditory leading visual stimuli by up to 21ms, to visual leading auditory stimuli by up to 150ms (Stone et al., 2001).

This begs the question of how faithful the merged percept should be with regard to the information conveyed through each sensory modality when there is a spatial or temporal disparity between visual and auditory sensory signals? Generally speaking, the visual system is superior to the auditory system in terms of spatial resolution, whilst the auditory system is superior to the visual system in terms of temporal resolution. Intersensory bias is usually governed by the dominant influence of one sensory modality over another, less reliable, sensory modality. In the spatial domain, visual stimuli can strongly bias the perceived location of a concurrently presented auditory stimulus, whereas the localised source of sound stimuli influences visual targets only very weakly, if at all (this is revisited below). On the other hand as the auditory system is specialised for the detection of rapid temporal variations in sound it can bias the perception of visual events in time, when the two are perceived concurrently.

Inherent in such a flexible system, tolerant to both temporal and spatial discrepancies between the visual and auditory percept, is the potential for errors. Many psychophysical demonstrations of multisensory integration depend on the perceptual outcome of these errors. Multisensory integration is frequently studied using intermodal conflict, where the information contained in one sensory modality conflicts with and perceptually modulates a stimulus in another modality. In these situations, the modality most reliable for accurately processing the feature about which there is disagreement across the two sensory streams, dominates or 'captures' the other sense to resolve the ambiguity, as in the ventriloquist illusion (described below in section 2.2.3). In some

circumstances, crossmodal intersensory modulation of conflicting or incongruent stimuli can result in a percept that is neither uniquely captured by the visual stimulus, nor the auditory stimulus, but is a synthesis of the two as in the McGurk effect (described below in section 2.2.6).

2.2.3 Vision tends to dominate audition in spatial tasks

It is well established that visual perception is more accurate than auditory perception in spatial tasks (Williams and Aitken, 1977). The art of ventriloquism has long been cited as the prime example of visual capture of acoustic spatial source (Bertelson and Aschersleben, 2003; Bertelson and Radeau, 1981; Howard and Templeton, 1966). When speech sounds are produced without visible movement of the ventriloquist's mouth, the illusion of a speaking dummy is created by substituting the appropriate visual transitions at the dummy's mouth in synchrony with the auditory transitions. This forces the observer to attribute the cause of the voice to the target in visual space where there is demonstrable visual-auditory temporal correspondence, despite the physical separation of the two sources. Thus, ventriloquism is an example of visual capture of an acoustic source, an illusion created by taking advantage of the auditory system's subordination to the visual system in terms of locating the source of an auditory cue. There is a window of synchrony between auditory and visual events that is crucial for the success of ventriloquism (Slutsky and Recanzone, 2001).

2.2.4 Audition tends to dominate vision in temporal tasks

In the temporal domain, there is strong evidence to suggest that the auditory system provides more reliable/accurate information than vision (Hirsh and Sherrick, 1961; Welch and Warren, 1980). In particular, temporal sequences are processed more efficiently in the auditory than visual domain (Cole et al., 1961; Klapper and Birch, 1971). Audition has been found to capture visual temporal perception, a phenomenon known in the early literature as 'auditory driving' (Gebhard and Mowbray, 1959; Myers et al., 1981; Shipley, 1964). Here, changes in the rate of a repetitive clicking sound induce corresponding changes in the perceived rate of a stationary, repetitive flashing light (Gebhard and Mowbray, 1959; Welch et al., 1986).

The superior temporal resolution of the auditory system (under normal circumstances) can go further to create 'illusory' visual events upon concurrent, rapid

presentation of sequential auditory stimuli (Sekuler et al., 1997; Shams et al., 2002). In this illusion, when two (or three) task-irrelevant acoustic beeps are presented simultaneously with a single light flash, subjects perceive two (or three) light flashes. Perception of the visual stimulus is thus influenced by the temporal sequence information contained within the accompanying auditory stimulus. This results in a crossmodally-integrated percept that shares features of both the visual and the auditory stimulus. Such is the propensity of the brain to pair together approximately coincident visual and auditory events according to the auditory system's more reliable (and therefore dominant) analysis of rapid temporal order.

Recently, a new form of crossmodal capture has been documented which is known as 'temporal ventriloquism' (Morein-Zamir et al., 2003). In this experimental manipulation, the perceived stimulus onset asynchrony between two sequentially presented light stimuli is enhanced by asynchronous, but not synchronous, presentation of two task-irrelevant auditory stimuli, when the first of these sound stimuli precedes the first light flash and the other follows the second light flash. This has the effect of perceptually separating the two visual stimuli in time and conveying improved accuracy in visual temporal order judgements. Conversely, when the auditory stimuli occur in between the two visual stimuli, the opposite effect is observed and the subjects demonstrate a diminished performance. This 'temporal capture' was strongly dependent on the second of the two temporally offset visual and auditory crossmodal 'pairings' and was abolished if they were presented synchronously.

2.2.5 Crossmodal capture or crossmodal averaging?

There are two frameworks in which the bimodal phenomena resulting from intermodal conflict can be considered. It can be viewed as a dominant sense, which due to its superior capabilities in determining spatial (vision) or temporal (auditory) information about some visual-auditory phenomenon, can bias the processing of information in the other modality. This view is known as the 'modality appropriateness hypothesis' (Welch and Warren, 1980). Several recent studies have invoked the view that such bimodal perceptual phenomena can be more accurately described as a visual-auditory crossmodal synthesis of two sources of information resulting in a weighted average according to signal strength and reliability (Battaglia et al., 2003; Heron et al., 2004). This view builds on visual-haptic crossmodal studies that have suggested similar

flexibility during perceptual integration of seen and touched sensory objects (Ernst and Banks, 2002; Ernst et al., 2000; van Beers et al., 2002).

Identity interactions, such as the 'McGurk effect', do not readily adhere to the basic concepts of visual spatial and auditory temporal dominance. The McGurk effect occurs with speech stimuli, whereupon hearing a /ba/ sound, whilst seeing the speaker form the shape of the syllable /ga/, a final merged perception of /da/ results (McGurk and MacDonald, 1976). As with all examples of intermodal conflict, the McGurk effect also depends on a degree of synchrony between visual and auditory stimuli, failing to occur when the visual-auditory asynchrony exceeds 200-300ms (Massaro et al., 1996; Munhall et al., 1996).

The flash lag effect, a phenomenon in which a sudden, stationary flash of light appears to be lagging behind a moving target, was considered a purely visual phenomenon for many years (e.g. Metzger, 1931; Nijhawan, 2002; Whitney, 2002). However, this same effect has recently been exhibited not only unimodally within the auditory system, but also crossmodally (Alais and Burr, 2003; Krekelberg, 2003). That a visual flash can appear to lag behind auditory motion through space (or behind a simple frequency sweep) and that an auditory beep can appear to lag behind visual motion, indicates the existence of reciprocal influences between visual and auditory sensory systems at some, as yet unknown, level of processing.

Other recent psychophysical studies have suggested that the rules governing crossmodal capture are yet more flexible than was previously assumed and do not always rely on visual spatial perceptual dominance and auditory temporal perceptual dominance (Bertelson and Aschersleben, 2003). When blurring is used to diminish the clarity of the visual cue, or visual spatial resolution is degraded in some other manner, ventriloquism fails, presumably because it no longer provides superior spatial information (Alais and Burr, 2004). This suggests that visual and auditory cues are integrated in a more probabilistic manner, whereby the spatial and temporal signal strength across the two modalities may be first taken into account, before crossmodal integration creates the final, merged, bimodal percept, according to the reliability of the available information in each sensory domain.

2.2.6 Summary

The models of entirely independent, separate, parallel processing for visual and auditory cues in dedicated unimodal cortical territories may have to be revised in light of the data presented above. The merged multisensory perception that results from crossmodal integration appears to be weighted according to the reliability of information conveyed by each sense, rather than a strict adherence to the doctrine of 'modality appropriateness' (Welch and Warren, 1980). The systems governing such interactions appear to operate in a very flexible manner, in that, as the information conveyed through one sense is compromised, a greater emphasis is placed on the information provided by the other senses (Alais and Burr, 2004). These integrative visual-auditory interactions appear to occur when information in one sensory modality is: 1) not sufficient to provide reliable information about the stimulus characteristics in question, and therefore 2) does not dominate the percept, because 3) there is equivalent stimulus information of higher fidelity available from the other sensory modality (Heron et al., 2004). In this situation the eventual holistic perception appears to comprise a weighted average of the two signals, according to relative signal coherence. Such a co-operative approach to sensory perception is a logical strategy to have evolved in mammals struggling for survival in constantly changing environments.

Nightfall, or the gloomy environments of caves or dense forests, can significantly compromise the usually reliable information conveyed by the visual system regarding the identity of specific objects or events. Similarly, periods of noise interference in the acoustic environment derived from rainstorms, rough seas or high winds, can all compromise the detection of spectrotemporal sound features that are usually reliably discerned by the auditory system. As the degree to which the efficiency of visual and auditory sensory function is reduced according to the time of day, immediate environment and prevailing weather conditions, it seems that multisensory perception can compensate by merging correlated sensory information according to its relative reliability at that time.

Next we turn to animal electrophysiology and anatomical literature in the search for clues to which brain structures might be the critical sites at which visual and auditory perception becomes integrated.

2.3 Anatomical & electrophysiological indications of visual-auditory integration

Most models of unimodal sensory brain organisation have generally described a hierarchical processing chain where the most simple, sensory features are extracted first followed by more and more complex features analysed at subsequent levels (e.g. Felleman and Van Essen, 1991; Hubel and Wiesel, 1962; Hubel and Wiesel, 1968; Rauschecker, 1997). Visual and auditory systems are largely segregated in terms of their neuroanatomical arrangement. This segregation remains more or less intact all the way from visual/auditory stimulus transduction at the retina/cochlea, during passage through the thalamus via the lateral/medial geniculate nuclei (LGN/MGN), to arrival at their respective primary sensory cortices within the calcarine sulcus located at the medial occipital cortex / at the superior temporal plane. The question that will be addressed in this section is simple: given that integration between these broadly separated sensory systems is so evident from the psychophysical literature, at what neuroanatomical level does this interaction occur? Studies of anatomical connectivity and electrophysiology are explored here in the search for evidence of sites and mechanisms of visual-auditory integration and sensitivity to crossmodal synchronicity.

First some terms will be defined. The term crossmodal integration (CMI) is used to imply anatomical convergence of projections from unimodal visual and auditory sensory cortices *at single cells* and the functional integration that results from this. In electrophysiology this is referred to as 'multisensory integration' and these terms have been used interchangeably when discussing this literature. CMI is distinct from multisensory convergence (MSC) in which separate visually responsive and acoustically responsive neuronal populations are in juxtaposition with one another, but in the absence of explicit evidence to suggest integration at individual neurons. Anatomical tracer studies can reveal sites of MSC, by demonstrating convergence of projections from separate unimodal cortical territories within the same anatomical area. Electrophysiology is required to establish whether these areas exhibit CMI, which can only be accurately determined by measuring responses to concurrent visual and auditory stimulation at single neurons.

2.3.1 Anatomical sites of multisensory convergence

Neuroanatomical and electrophysiological recordings in animals throughout the latter half of the last century have revealed a myriad of MSC and CMI sites. At the subcortical level, multisensory integration has been demonstrated in the reticular activation system (Amassian and Devito, 1954; Bell et al., 1964; Yen and Blum, 1984), pooling input from the various sensory exteroceptors. This drives global arousal via diffuse projections that exert a profound influence on activation across the whole cortical mantle. Subcortical multisensory neurons have also been recorded at the locus coeruleus (Grant et al., 1988), the external nucleus of the inferior colliculus (Itaya and Van Hoesen, 1982; Tawil et al., 1983), the superior colliculus (Gordon, 1973; King and Palmer, 1985; Meredith and Stein, 1983; Meredith and Stein, 1985; Meredith and Stein, 1986; Stein, 1984; Stein and Arigbede, 1972; Wickelgren, 1971) and cerebellum (Azizi and Woodward, 1990; Caan et al., 1976; Chapman et al., 1986; Freeman, 1970). In the thalamus, the segregation of visual and auditory primary projection pathways in the lateral and medial geniculate nuclei, respectively, is accompanied by multisensory convergence in posterior and lateral thalamic regions (Chalupa and Fish, 1978; Rasmussen et al., 1984). Visual and auditory stimuli can also modulate activity in the basal ganglia, an interrelated group of structures intermediate between the brainstem and cortex, that play an essential role in coordinating movement (Hikosaka et al., 1989; Hikosaka and Wurtz, 1983; Strecker and Jacobs, 1985). In the neocortex, primate superior temporal, inferior parietal and orbitofrontal cortices are also well endowed with multisensory neurons (Benevento et al., 1977; Bruce et al., 1981; Ito, 1982; Loe and Benevento, 1969; Stein et al., 1993; Vaadia et al., 1986; Watanabe and Iwai, 1991).

2.3.2 Crossmodal integration at the superior colliculus

No description of neuronal visual-auditory integration would be complete without giving a special account of the superior colliculus. Considered something of a gold standard of multimodal integration, this midbrain structure contains overlapping maps of visual, auditory and tactile space, exploiting special intrinsic properties that amplify the efferent signal when concurrent stimulation occurs in two or more sensory modalities in the preferred region of space.

The superior colliculus was originally considered a visuomotor structure based on the construction of a crude movement map inferred from gross electrical stimulations

(Adamuk, 1870), and the visuotopic electrophysiological responses in superficial layers to visual stimulation of the retina (Apter, 1945). Anatomical studies of the forties and fifties soon established the existence of substantial non-visual afferents into its deeper layers (Anderson and Berry, 1959; Marburg and Warner, 1947; Mehler et al., 1960; Poirier and Bertrand, 1955) and by the sixties, electrophysiological papers surfaced documenting neuronal responses to visual and auditory stimuli (e.g. Bell et al., 1964). Of particular note was the absence of any tonotopic organisation in the auditory responses of deep layer superior colliculus neurons, which is a defining characteristic of the thalamocortical auditory system (Reale and Imig, 1980). These were instead particularly responsive to complex stimuli, such as jangling keys, hisses and hand claps, compared to pure tones (Horn and Hill, 1966; Stein and Arigbede, 1972; Wickelgren, 1971). The fact that these neurons also responded best to moving stimuli (Gordon, 1973; Rauschecker and Harris, 1989), suggested that they were specialised not for stimulus identification, but for stimulus localisation.

It is now broadly accepted that the superior colliculus is optimised for rapid orientation of the sensory organs towards simultaneous stimulation in overlapping visual, auditory and tactile space. Certain principles governing CMI at the neuronal level, enabling a distinction from MSC, have emerged from this work (Meredith and Stein, 1986). These 'rules' state that CMI involves: (a) *superadditive* responses to contiguous bimodal stimulation, (b) *response depression* to non-contiguous bimodal stimulation and (c) observance of the law of *inverse effectiveness*. *Superadditive* responses are found to be widespread among cells of the deep laminae and are strongly indicative of crossmodal integration. This form of response enhancement typically involves a significant increase in the number of discharges evoked by spatially and temporally coincident bimodal stimuli that is multiplicative, rather than summative, when compared to the respective unimodal responses. *Response depression* is characterized by a significant decrease in the discharges elicited in response to spatially or temporally offset, bimodally presented visual and auditory cues, as compared to the unimodal responses. Finally the law of *inverse effectiveness* stems from the observation that the response enhancement measured during bimodal stimulation was generally inversely related to the strength of the responses evoked by presenting each unimodal stimulus alone. In other words the, 'response amplification was greatest when responses evoked by individual stimuli were weakest' (Meredith and Stein, 1986).

Recently, the crossmodal integrative functions of the superior colliculus have been shown to depend critically upon influences from frontal brain regions (Wallace and Stein, 2000). Temporary cryogenic deactivation of cat anterior ectosylvian sulcus and rostral lateral sulcus totally eliminated response enhancement in deep layer neurons to appropriate stimuli, suggesting that cortical brain functions play a greater part in mesencephalic crossmodal interactions than previously imagined.

2.3.3 Crossmodal integration in the neocortex

Early anatomical studies indicated convergent monosynaptic cortico-cortical inputs from visual, auditory and somatic association areas onto specific regions of the superior temporal sulcus (Gross et al., 1967; Merzenich and Brugge, 1973), insula (Mesulam and Mufson, 1982) and polysynaptic convergence at various frontal regions such as the orbitofrontal cortex (OFC), via neurons of the temporal pole (Chavis and Pandya, 1976; Jones and Powell, 1970; Kuypers et al., 1965; Pandya and Kuypers, 1969; Seltzer and Pandya, 1978). Multimodal anatomical projections have also been demonstrated at inferior parietal sites, predominantly from visual and somatosensory sources (Seltzer and Pandya, 1980). More recently, substantial input into the ventral intraparietal area has been identified from extrastriate visual, somatosensory, motor, vestibular, polysensory cortical areas and also from auditory areas (Lewis and Van Essen, 2000).

Electrophysiological investigations in macaque have characterised the multisensory responses of neurons in the superior temporal sulcus (Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988), inferior temporal areas (Desimone and Gross, 1979) and lateral OFC (Benevento et al., 1977), in an effort to ascertain whether bimodal neurons could be identified. These studies primarily investigated responses of cells to *unimodal* visual, auditory and somatosensory stimulation in order to accurately define the modality and feature specificity, rather than the response interactions during concurrent, *bimodally* presented stimuli (e.g. Bruce et al., 1981). These have been useful in characterising the incidence of bimodal cells, as discerned from their responses to both visual and auditory unimodal stimulation, with estimates ranging from 21 - 36% of cells in the anterior STS and 43% in the lateral OFC (Benevento et al., 1977; Bruce et al., 1981). These studies were also useful in their primary aim of describing stimulus preferences. A common finding was that the majority of neurons in the superior

temporal polysensory area were particularly responsive to moving stimuli in large receptive fields (RFs) approaching the size of the entire visual field, and were often sensitive to the specific direction of motion (e.g. Bruce et al., 1981). It was also noted that visual and auditory RFs could be overlapping, or complementary, i.e. auditory RFs were juxtaposed to the periphery of the visual RFs (Hikosaka et al., 1988). In many respects bimodal cells were found to be very unselective for specific stimulus features, demonstrating similar responses to a variety of visual stimulus sizes, shapes, orientations or contrasts and whether auditory stimuli were clicks, tones or vocalisations.

In the few studies that investigated responses to combined, bimodal stimulation, the most prominent visual-auditory interaction was response depression (e.g. Benevento et al., 1977). Specifically, it was observed that auditory stimuli at a specific frequency could have a potent inhibitory effect on many bimodal neurons, in many cases negating the excitation caused by visual stimuli. This effect was predominantly found in OFC, but also in the superior temporal polysensory area. Bimodal facilitation was also exhibited at OFC cells when the temporal onsets of the visual and auditory stimuli were manipulated appropriately. This led to the proposition that the early hyperpolarization or subsequent excitatory discharge in response to the auditory stimulus, when appropriately phased with respect to the depolarisation effect of the light stimulus, could produce either suppression or facilitation of the spike discharge (Benevento et al., 1977). This suggests the possibility that in this area such bimodal cells may not be tuned so much to the specific visual and auditory stimulus characteristics, but to the specific temporal relationship between them. In all these studies little or no superadditive activity was detected during concurrent visual and auditory stimulation. Likely reasons for this are evaluated in the discussion (section 4.3.1).

Superadditivity has since been documented in a new class of bimodally responsive neurons recently isolated in monkey ventral premotor cortex (Kohler et al., 2002). 'Mirror neurons' were first described based on measured discharges both when a monkey performed a specific action and when it observed another individual performing that same action (Gallese et al., 1996; Rizzolatti et al., 1996). A more recent study claims to have measured neuronal discharges in a class of neuron they have dubbed 'audiovisual mirror neurons' in response to the sight or sound of an action, as well as in response to the monkey performing that action themselves (Kohler et al., 2002). These neurons fell into two categories: half responded equally under unimodal or

bimodal stimulus conditions, but 8/22 neurons exhibited characteristics of CMI whereby the strongest response was observed when the sight and the sound of the action were presented together, compared to apart. Five of these cells demonstrated additive bimodal responses (VA=V+A), whilst the other three were unresponsive to visual stimulation, but bimodal visual and auditory stimulation resulted in a higher rate of discharge than unimodal auditory stimulation (VA>{V}+A) i.e. superadditivity. Receiver operator characteristic analysis revealed that if the monkey based its ability to differentiate between two actions on the activity of one of these cells, their performance would be approximately 90% accurate using unimodal cues, but 97% correct based on the bimodally combined visual and auditory cues.

2.3.4 Summary

The electrophysiology literature offers evidence for neurons exhibiting responses to separate, and occasionally combined, visual and auditory stimulation at several subcortical areas, but also at several neocortical sites within the temporal, frontal and parietal lobes. The various psychophysical and behavioural outcomes of merging visual-auditory stimuli during human bimodal perception (section 2.2) may be critically supported by crossmodal integration of sensory information at any of these sites. Furthermore the tuning of lateral orbitofrontal cortex neurons to visual-auditory temporal offsets suggest that this region may be a candidate functional unit involved in regulating crossmodal integration. The neurology literature is considered next in the search for further candidate neuroanatomical areas that may be implicated in visual-auditory perceptual integration.

2.4 The neurology of visual-auditory perception

Neurological conditions that disrupt normal visual-auditory sensory perception in human patients may indicate brain regions involved in multisensory integration. In the first part of this section the neural correlates of synaesthesia are considered. Then the symptoms of neglect and the brain lesions that induce this condition are described. The neuroanatomical loci implicated in these conditions may provide *a priori* hypotheses guiding the search for neural correlates of visual-auditory integration.

In developmental psychology there are two opposing perspectives regarding the state of sensory integration at birth. The so-called law of "specific nerve (read 'sensory') energies" (Müller, 1826) has been very influential in sensory research, without doubt assisting the development of a better understanding of individual sensory systems, but resulting also in scientific models that strictly partition the processes supporting perception through different modalities. Regarding the ontology of multimodal perception, it is not clear whether infants begin life with a primitive unity of the senses that becomes differentiated through development (Ryan, 1940; Werner, 1934), or whether the senses are strictly separated at birth with connections forged through experience of joint occurrence (Gibson, 1966; Piaget, 1953). Both possibilities being equal, it is certainly worth considering human cases of fused sensory experience such as synaesthesia. In the most common form of synaesthesia the perception of a certain sound form is accompanied by the sensation of a particular colour, but in the absence of any visual stimulation. Synaesthesia could therefore be viewed as but one extreme on a continuum of intersensory bias (see section 2.2). It can be considered a special, but aberrant, case of crossmodal perception, the neural mechanisms of which may illuminate important areas responsible also involved in normal crossmodal perception.

Conversely, when damage to specific brain regions results in perceptual poverty affecting both the visual and auditory senses, as in neglect, it implies that these areas are vital for conscious perception of visual and auditory stimuli. If these brain areas are involved in both visual and auditory sensory processing then it is conceivable that they may have a role to play in crossmodal integration.

2.4.1 Synaesthesia

A systematic study of synaesthetic experience was conducted towards the end of the nineteenth century (Galton, 1883). The most common manifestation of clinically presented synaesthesia is in the visual-auditory domain, perhaps indicating that the visual and auditory perceptual systems may be particularly closely related. The most prevalent form of this condition is chromatic-lexical, or colour-word, synaesthesia, where a spoken (auditory) word gives rise to the visual perception of colour. However, upon *hearing* a spoken word, it is the *visual form* of the first letter of that word and NOT the *auditory phonology* of the first syllable of the word that governs which *visual* colour is perceived. For instance, if the synaesthete were to hear the words 'photograph' and 'fish', they perceive a different colour for each, but the words 'photograph' and 'police' would both elicit perception of the same colour (Paulesu et al., 1995). This may be interpreted as endorsement of the scheme advocated by Piaget and Gibson (see preamble to this section) indicating that intersensory associations are generated during development and that this is reflected in synaesthesia as a result of the formation of inappropriate crossmodal links during the acquisition of language skills and in particular during early vocabulary building.

In the first imaging study undertaken to assess the neural correlates of synaesthesia, Paulesu and colleagues (Paulesu et al., 1995) scanned six synaesthetes six and six normal control subjects using positron emission tomography (PET), who were simply instructed to tap their left index finger each time they heard a tone or a spoken word. Synaesthetes, but not controls, demonstrated increased activity during auditory word processing in the left posterior inferior temporal (PIT) cortex, bilateral parietal-occipital junction, right inferior / middle frontal gyrus, insula and superior temporal gyrus, and deactivations at the left lingual gyrus and insula (Table 1, left column). One or all of these regions could potentially be involved in crossmodal integration given that this activity occurred during acoustically induced bimodal perception (of colour/sound) and was significantly different from that of non-synaesthetes in whom the crossmodal perceptual phenomenon was absent. Significant activations were not observed in synaesthetes in the classical visual colour pathway i.e. visual areas V1, V2 and V4. It was noted that colour perception resulting from auditory perception of spoken words in synaesthetic experience is not, in fact, generally observed to be attached to a specific object, nor even any particular quadrant of space (Baron-Cohen et al., 1993). It was

proposed that the absence of colour pathway activation during synaesthetic experience fitted well with the observed degradation of retinotopic organisation found in the ventral visual areas anterior to V4 (Komatsu et al., 1992), which is in turn far less precise than that of V1 (Essen and Zeki, 1978). However, a more recent functional Magnetic Resonance Imaging (fMRI) study of this form of synaesthesia overturned this notion, finding that the word sound-induced colour perception *did* induce activation in area V4 and V4α (Nunn et al., 2002). This study also implicated activations in several new regions during merged visual-auditory perception in synaesthetes that were absent in normal subjects. These include voxels within the right medial and superior frontal gyri, several areas within left posterior cingulate, the right claustrum and left inferior parietal lobule (Table 1, right column).

These non-invasive imaging studies of synaesthetic experience have revealed significantly active voxels at a variety of different neuroanatomical locations, but with considerable differences in the distribution of these activations (see Table 1). However among these different sites surely lies the true neural correlates of synaesthetic experience. Such regions might also be involved in integrated bimodal perception in normal subjects during concurrent visual and auditory stimulation.

Paulescu *et al.* (1995)	**Nunn *et al.* (2002)**
Posterior inf. temporal cortex (L+: -54,-42,-16)	V4 (L+: -33 -66 -13)
	posterior cingulate (L+: -19,-55,15; -16,-60,9)
superior temporal gyrus (R+: 54,-10,4)	
lingual gyrus(L-: -8,-66,0)	
parietal-occipital (L+: -30,-62,40; R+: 26,-64,40*)	inf. parietal lobule (L+: -40,-58,35; -27,-27,53)
insula (R+: 40,8,0; L-: -28,-2,-4)	claustrum (R+: 33,-6,4)
inferior frontal gyrus (R+: 36,8,28)	
middle frontal gyrus (R+: 30,50,8)	
	superior frontal gyri (R+: 22,27,48)
	medial frontal gyri (R+: 7,42,-7)

Table 1: Neuroanatomical areas implicated in functional imaging studies of synaesthesia. The PET study of Paulescu and colleagues (1995) and the fMRI study of Nunn et al. (2002) both scanned the brains of subjects during synaesthetic experience of colours during unimodal auditory stimulation. They revealed distributed activations in several different brain areas (co-ordinates in Talairach stereotaxic space) in synaesthetes, but not control subjects and there was little overlap between the findings of the two studies. ('L' = left side; 'R' = right side; '+' = activation; '-' = deactivation; '*' = see results section 3.3.2)

2.4.2 Neglect

Neglect is a neurological condition that results from trauma affecting right lateralised brain regions and has been known for some time to lead to disrupted perception of contralateral visual space (Lawson, 1962). More recently, clinical evidence has since been published to indicate that not only can neglect also disrupt non-spatial aspects of visual perception (Husain et al., 1997), but that it also induces a variety of auditory deficits (Bisiach et al., 1984; Heilman and Valenstein, 1972). As neglect seems to affect both visual *and* auditory perception, it is likely that the convergence of bimodal information may occur within the anatomical regions compromised by neglect lesions and it is therefore feasible that such areas may also be responsible for functional integration of correlated visual and auditory cues. Motivated by this possibility, the following section will briefly describe the spatial and non-spatial aspects of visual and auditory sensory deficits presented by neglect patients and outline the key regions of overlap for the associated lesions.

Visual neglect

Unilateral spatial neglect is a common neurological syndrome in which visual awareness of contralesional space is disrupted. Patients suffering from neglect typically fail to report, respond or orient towards visual stimuli presented in the contralesional hemifield (review: Marshall and Fink, 2001). The perceptual manifestations of neglect were for many years considered to involve only *visual* awareness of *contralesional space*, but upon closer inspection there are many exceptions to this view.

The modern neglect literature includes examples where attention *can* be successfully directed to targets in contralesional space. For instance, so long as the task involved detection of a specific target colour, upon presentation of two adjacent visual objects, neglect patients could selectively identify both the contralesional and the ipsilesional object (Duncan et al., 1999). Conversely, neglect patients can be equally poor in their ability to detect certain stimuli on both sides. For instance, during a selective attention and a multiple object tracking tasks, deficits were observed in the perception of apparent motion, which equally affected the ipsi- and contralateral hemifields (Battelli et al., 2001).

Neither is neglect of contralesional *space* the only visual deficit in neglect patients (reviews: Husain and Rorden, 2003; Robertson, 2001). In one particular study, the ability of neglect patients to detect two target letters in a rapid serial presentation paradigm was tested and their performance compared to that of both stroke patients who did not exhibit neglect and to normal, healthy volunteers (Husain et al., 1997). We are all, as a natural consequence of our finite attentional resources, incapable of identifying two consecutive targets within ~400ms, a phenomenon known as the attentional blink (Broadbent and Broadbent, 1987; Duncan et al., 1994; Raymond et al., 1992; Shapiro et al., 1994). Neglect patients, however, demonstrated abnormally severe and protracted attentional blink of three times (~1440ms) that of normal and non-neglect lesion control subjects (Husain et al., 1997). Thus, for the first time, neglect was shown to affect the ability to detect different visual objects in time, as well as space.

Recently, Pavani and colleagues (Pavani et al., 2003) have reviewed a host of auditory deficits that accompany the visual deficits and have also identified a statistically significant correlation in the severity of visual and auditory neglect symptoms (Pavani et al., 2004). Before considering the neuroanatomy of neglect the auditory symptoms of neglect will be briefly outlined.

Auditory neglect

Anecdotal clinical evidence of auditory neglect was first reported by Denny-Brown and colleagues (Denny-Brown et al., 1952). It was observed that, despite having good hearing in both ears, patients would report the detection of a sound presented to the left ear as having come from the right (Denny-Brown et al., 1952; Diamond and Bender, 1965), or that patients failed to respond when verbally addressed from contralesional space (Battersby et al., 1956; Heilman and Valenstein, 1972). Not until three decades later was a thorough investigation undertaken by Bisiach and colleagues (Bisiach et al., 1984), which found a systematic shift in the localisation of dichotically presented sounds, towards ipsilesional space, in 107 neglect patients. They hypothesised that, as the ability to localise auditory as well as visual stimuli in space were both compromised, lesions that induce neglect might actually disrupt the internal representation of egocentric, modality-independent (amodal) space.

In recent times, several rigorous analyses of auditory spatial neglect have supported this view, finding that there is an overall trend for right-sided parietal lesions to bias sound localisation to the right, presumably due to the compromised ability to

process stimuli in the contralesional left-hemispace. Several studies have demonstrated this by asking patients to point to the sound source (Altman et al., 1979; Pavani et al., 2003; Vallar et al., 1995), adjust the panning of a continuous tone until it is perceived as central (Kerkhoff et al., 1999; Tanaka et al., 1999), or declare whether two consecutively presented tones were from the same or different location (Pavani et al., 2001; Tanaka et al., 1999). However other studies have indicated that there are exceptions to these findings and that the phenomenon is more complex than this simple rule of thumb (e.g. Cusack et al., 2001; Pinek et al., 1989; Ruff et al., 1981).

Visual tests of neglect usually involve several simultaneously presented objects, whilst auditory tests tend to use presentation of a single sound in isolation. To specifically address these inequalities, several dichotic listening tasks have been developed where two different sounds are presented binaurally to the two ears. This rendered patients unable to identify stimuli at the contralesional ear, or resulted in failure to detect them at all (Bellmann et al., 2001; Deouell and Soroker, 2000; Soroker et al., 1997).

Neglect lesion anatomy

Neglect is typically caused by major strokes affecting the vascular territory of either the middle or posterior cerebral arteries, often affecting widely distributed regions of the cortical mantle (Mort et al., 2003). A particularly high incidence of neglect results from right-lateralised lesions affecting the border between the temporal and parietal lobes; an area known as the temporo-parietal junction (TPJ: figure 4) (Heilman et al., 1983; Vallar, 2001; Vallar and Perani, 1986). However, the variability in brain areas compromised by stroke and resulting in neglect recently gave rise to an intense debate, sparked by Karnath and colleagues (Karnath et al., 2001), who controversially suggested that it was damage to the mid-portion of superior temporal gyrus that was critical to the induction of neglect. This went against the broadly accepted view that neglect primarily results from lesions affecting the angular gyrus in the posterior part of the inferior parietal lobule (Mort et al., 2003). Assessment of the available data suggests that this disagreement may be due to: (a) differences in inclusion criteria in terms of size and location of lesions considered in the analysis, (b) differences in the analysis methodology, or (c) the most parsimonious conclusion: that both parties are correct in that injury to either area induces neglect, perhaps via disruption to different mechanisms. In fact, perisylvian lesions separated from the TPJ by the entire length of

the Sylvian sulcus can also give rise to neglect symptoms (review: Husain and Rorden, 2003). For instance, lesions of both the inferior and middle frontal gyri, dorsally adjacent to the anterior terminus of the Sylvian fissure (figure 4), can also induce visual contralateral spatial neglect (Husain and Kennard, 1996; Husain et al., 1997) and it would be very interesting to know whether such frontal lesions might induce auditory and non-spatial visual symptoms of neglect. If such lesions did disrupt processing and perception of audition as well as vision, then these frontal sites could join the TPJ as candidate sites at which visual and auditory sensory stimuli might be brought together into merged bimodal perception.

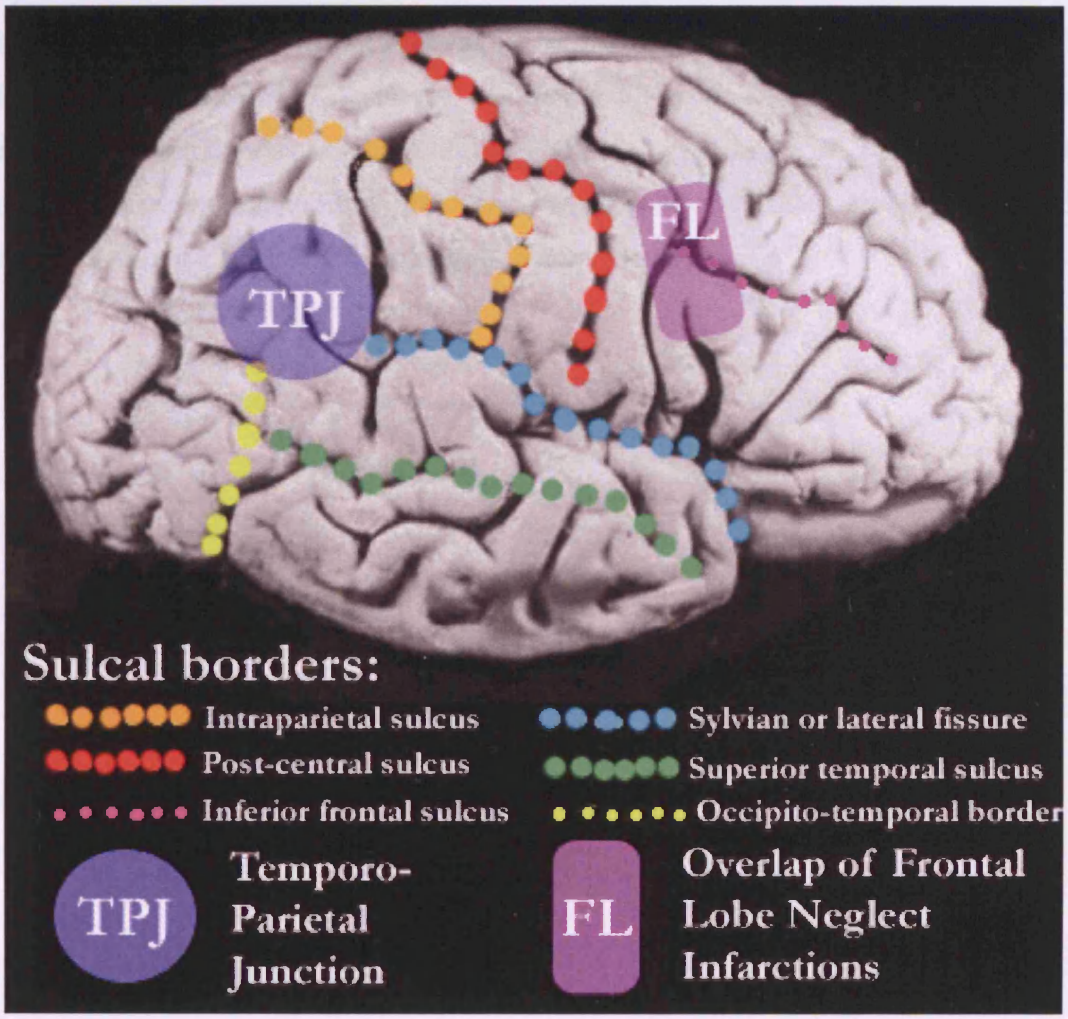


Figure 4: Neuroanatomical lesion sites known to induce neglect in man. This photograph of the right lateral view of the human cerebrum is annotated with the areas of overlap for posterior and anterior perisylvian lesions known to induce neglect in man. The posterior region is known as the temporo-parietal junction (TPJ) and straddles the anatomical territory of the ventral inferior parietal lobe and posterodorsal superior temporal gyrus. The anterior region is located in the frontal lobe (FL), including the posterior territory of both middle and inferior frontal gyri.

2.4.3 Summary

The synaesthesia and neglect literature indicate several neuroanatomical sites at which visual and auditory sensory information might be brought together and perhaps even result in the integration of the two senses into the subjective impression of sensory unity. This statement is based on the observation that, in neglect, lesions of the TPJ can compromise both visual *and* auditory awareness. The function of this area has traditionally been understood to revolve around mechanisms of spatial attention (Driver and Mattingley, 1998), but this does not preclude the possibility that the TPJ may also play a role in crossmodal integration. Visual perception induced by auditory sensory stimulation in synaesthetic individuals is accompanied by activity at a variety of additional anatomical loci, such as the superior temporal gyrus, parieto-occipital junction and frontal, posterior cingulate and inferotemporal cortices. These (tentatively) suggest further neuroanatomical sites of multisensory convergence at which crossmodal integration might also occur in the normal human brain.

2.5 Functional imaging of human visual-auditory perception

Having consulted both animal and neurology literature for possible neuroanatomical loci at which visual and auditory information might converge, the relevant functional imaging evidence is briefly considered in this section. Just as electrophysiological observations of visual-auditory interactions can be used to identify sites of crossmodal integration in the brains of experimental animals (King and Palmer, 1985; Meredith et al., 1987; Meredith and Stein, 1986), the neural correlates of integrated bimodal perception can be investigated in man using non-invasive imaging techniques. However, there are considerable differences between the two approaches that have to be borne in mind. Firstly, whilst electrophysiology directly measures the electrical activity within a single neuron, measurements in functional imaging indirectly reflect the combined activity of many thousands of neurons. Secondly, as electrophysiological recording can only measure activity in a few neurons at a time, it is tightly constrained with regard to *a priori* anatomical hypotheses. In other words, although electrophysiology permits accurate characterisation of neuronal activity profiles, it can only do so at the anatomical location at which the microelectrode has been inserted. On the other hand, although the signal measured in non-invasive techniques reflects either the metabolic demands (e.g. fMRI, PET), or electromagnetic activity (e.g. EEG, MEG) of many thousands of different neurons, they have the advantage of sampling neuronal activity across the whole brain, without having to rely entirely on *a priori* hypotheses regarding where the interesting responses might be found. Thus, a survey of the functional imaging literature can both confirm in human subjects the involvement of sites of neuronal integration detailed in the animal literature and potentially uncover previously unknown sites of multisensory convergence, as yet unexploited by animal electrophysiology.

2.5.1 Functional imaging of multisensory convergence or crossmodal integration?

As stressed by the Calvert group (e.g. King and Calvert, 2001), in order to differentiate between integration and convergence in functional imaging, it is not enough to simply examine the overlap or conjunction of areas activated by two or more different unimodal stimuli. As several animal studies have shown (e.g. Benevento et al.,

1977; Bruce et al., 1981; Desimone and Gross, 1979; Hikosaka et al., 1988), multisensory areas of the cortical mantle tend to contain mixtures of neurons responsive only to visual *or* auditory stimuli (unimodal cells), as well as those driven by *both* modalities (bimodal). Consequently, as a single voxel in an imaging experiment samples the collective metabolic demands of many thousands of neurons, any response to stimuli in more than one modality may simply reflect the presence of different populations of unimodal neurons in the same area. This has largely been overlooked by the vast majority of functional imaging papers to date and in this section, a critical review of this literature will outline the common errors made by researchers in this field, in order to support the methodology that was implemented in the studies comprising this thesis.

To infer crossmodal integration (CMI) within a voxel, one must demonstrate that the activation in response to *concurrent*, bimodal, visual and auditory stimulation differs significantly from the sum of activity from *separate*, unimodal, exposure to each of the visual and auditory stimuli in isolation. The vast majority of imaging experiments investigating visual and auditory stimulation of the human brain utilise *either* unimodal *or* bimodal paradigms. Whether regional activations found in such studies reflect crossmodal integration or multimodal convergence cannot be determined unless *both* unimodal and bimodal responses are collected. Most previous studies are therefore limited to declarations of whether a region was more active during: 1) stimulation in one sensory modality versus another, 2) under one set of bimodal conditions in contrast to another, or 3) when the task varies and the bimodal stimuli remain constant, i.e. under one set of behavioural constraints versus another.

Several fMRI studies have tried to interpret common regional activations that result from unimodal stimulation in different sensory modalities as crossmodal processing. For instance, Lewis and colleagues (Lewis et al., 2000) were the first to use functional imaging to investigate the crossmodal perception of motion. They found that separate unimodal visual and auditory motion perception tasks conjointly activated lateral parietal, lateral frontal, anterior midline and anterior insula cortex. However, stimuli were never presented in combination and so regions of CMI could not be defined. They can only claim to have found evidence of multisensory convergence (MSC). Nonetheless, they attempted to infer CMI using a crossmodal speed comparison paradigm in which motion presented in one modality was immediately followed by and

compared against motion in the other modality. This task invoked activity in the same brain regions, with the exception of the lateral frontal activation.

Another recent human fMRI experiment extended these findings to include motion perceived through somatosensory cues in addition to visual and auditory motion. This work provided evidence to support a hypothesis, based on monkey electrophysiological studies, which proposed that amodal motion processing occurs in discrete regions of parietal and frontal cortex (Bremmer et al., 2001). They contrasted brain activity under unimodal presentation of either a moving visual random dot pattern (V), illusory auditory motion using binaural beats (A), or somatosensorily-derived motion from the passage of air across the face (T), against a motionless control for each. They used three separate conjunction analyses between each possible pair of combinations of sensory modality versus its control (V-control x A-control; V-control x T-control; A-control x T-control) to infer common activation in the deep inferior parietal sulcus (IPS), ventral premotor cortex and lateral inferior postcentral cortex. Here they can claim to have identified a region of MSC, but once again in the absence of bimodal stimulus conditions, it was not possible to infer CMI from these findings.

2.5.2 Contributions from electroencephalography

Human electroencephalographic (EEG) studies of bimodal versus unimodal stimulus processing traditionally produce averaged difference waves. These result from contrasting the event related potentials (ERPs) arising from bimodal (VA) compared to unimodal (V or A) stimulation conditions, using the formula: VA - (V + A). ERPs can be compared in this way having been collected during simple object detection tasks in which an object can be discriminated based upon its visually discerned shape, acoustically determined tone frequency, or both in tandem (Giard and Peronnet, 1999). This enables the successive activation of different brain regions to be charted with great temporal accuracy, but always with a degree of uncertainty about the source producing these activations. Despite lacking the spatial resolution of imaging techniques like fMRI, EEG has the advantage of excellent temporal resolution. Statistical assessment of the most likely cerebral origins for the activity detected at the scalp surface can be accomplished through the use of dipole-modelling techniques such as brain electrical source algorithms (BESA). However, the lack of a definitive, unique solution to this inverse problem casts a shadow of uncertainty over the neuroanatomical source of EEG

recordings, which is often considered to be an unsatisfactory compromise. With these shortcomings acknowledged, in the Giraud study the expected improved detection accuracy and speed, was accompanied by new visual activations as early as 40ms over occipitoparietal cortex, further intersensory modulations at the auditory cortex from 90-110ms and later still sites of crossmodal integration peaking at 140-165ms over right lateralised anterior perisylvian cortex. However, such findings have been dismissed as artifactual because of significant caveats stemming from the subtraction of anticipatory slow waves within the unimodal ERPs twice from the single bimodal ERP (Teder-Salejarvi et al., 2002).

2.5.3 Attention to visual and auditory stimuli

Robert Downar and colleagues have conducted a series of fMRI experiments where changes in visual (V), auditory (A) and tactile (T) stimuli were either passively (Downar et al., 2000; Downar et al., 2002), or actively (Downar et al., 2001) detected. These studies all highlighted the importance of a region of cortex at the posterior extent of the right Sylvian fissure, the temporal-parietal junction (TPJ), previously discussed in the neglect review (section 2.4.2). In the first experiment (Downar et al., 2000), all stimuli were presented simultaneously and perceived passively, but changes would occur in only one modality at a time. This revealed a large modality-independent, change-sensitive region of activation in the right TPJ and smaller activations in left TPJ, right insula, right inferior post-central sulcus & left supplementary motor area/ cingulate motor area (SMA/CMA). In the second study (Downar et al., 2001), only visual and auditory stimuli were used, but subjects were instructed to attend to changes in only one modality and to ignore the others. Bilateral TPJ, left anterior insula, left precuneus and left anterior cingulate were observed to be more active when the modality was attended than unattended, whilst right insula, right inferior central sulcus and left SMA/CMA were equally active whether stimuli were attended or unattended. Returning once more to the passive-attention paradigm (Downar et al., 2002), visual, auditory and tactile salience was manipulated by introducing unfamiliar stimulus changes amongst familiar stimulus changes. Elevated activity was observed in the right TPJ, right anterior insula, right inferior central sulcus and left anterior cingulate in response to novel, unfamiliar stimulus changes, in comparison to familiar stimulus changes.

Consequently, Downar and colleagues have proposed that the combined evidence of these three experiments indicates that the right TPJ is sensitive to stimulus salience, irrespective of the sensory modality and regardless of whether its salience is inherent in the stimulus features ('bottom-up'), or cognitively generated according to prior instructions ('top-down').

2.5.4 Detection of visual-auditory synchrony

Human psychophysics, as delineated in section 2.2, has demonstrated that a much greater reliance is placed on temporal contiguity between bimodal visual and auditory cues than spatial contiguity (e.g. the ventriloquist illusion). Two separate human imaging experiments have demonstrated that temporal synchronicity between discrete pairs (Bushara et al., 2001) and trains of visual and auditory stimuli (Calvert et al., 2001) can be manipulated to reveal functional brain units that are differentially responsive to visual-auditory synchrony versus asynchrony. These studies are highly relevant to the issue of determining the site of neural mechanisms that establish whether bimodal stimuli are suitable for crossmodal integration or not.

Bushara and colleagues (Bushara et al., 2001) undertook a PET study investigating the effect of systematically varying the interval between the onset of single, brief, transient visual and auditory stimuli. In a subtraction analysis contrasting responses during exposure to asynchronous versus synchronous events, significant activations were found in the right insula, right inferior frontal gyrus, right inferior parietal lobe (IPL) and left cerebellum, whether visual stimuli preceded or followed auditory stimuli. The superior colliculus, left insula, right precuneus, right posterior thalamus and right PFC were subsequently implicated in an inter-regional covariance analysis that identified voxels with activity profiles that co-varied with that of the right insula. The right insula was of particular interest because its response magnitude varied parametrically with the degree of asynchrony.

In an fMRI experiment (Calvert et al., 2001), brain areas sensitive to synchrony detection were scrutinised in a passive attention paradigm, which included periods of both unimodal and bimodal stimulation. A rapidly alternating (8Hz), black and white checkerboard stimulus was used, reversals of which were synchronous with the onset of brief white noise stimuli in one session and asynchronous in another session. They observed that the superior colliculus, insula, STS and IPL were differentially activated

across these two conditions, with greater activity when the stimuli were presented synchronously than when a variable temporal offset was introduced. Furthermore, they also detected response depression in these regions during exposure to asynchronous stimuli, which they put forward as evidence of crossmodal integration as per electrophysiologically defined criteria (section 2.3.2). Other regions demonstrating classical crossmodal responses were the right-lateralised IPS, superior / ventromedial / anterior-posterior lateral sulcus frontal regions and left lateralised superior occipital gyrus. Several areas only demonstrated superadditive responses under synchronous conditions, in the absence of response depression during asynchronous stimulation, including both anterior and posterior cingulate cortex, precentral and middle frontal gyri on the right side, with parieto-occipital junction, IPL and lingual gyri on the left.

2.5.5 Recognition of visual and auditory objects

In an fMRI study comparing congruent to incongruent bimodal stimuli, crossmodal sensory processing was identified in the anterior cingulate and medial prefrontal cortex (Laurienti et al., 2003). It was suggest that these regions are responsible for sorting and coupling information derived from the same event based on the semantic or contextual congruence that exists between them. Alternating 30s blocks of visual-auditory stimulus pairs were presented in which the sound either semantically matched a line drawing of an object, or was mismatched in terms of its identity.

In another recent publication, Beauchamp and colleagues (Beauchamp et al., 2004b) conducted a series of experiments investigating visual and auditory objects processing. The combined evidence from three separate experiments implicated a region involved in visual-auditory object integration that varied considerably in location from subject to subject, between the posterior region of the superior temporal sulcus and middle temporal gyrus (pSTS/MTG).

Both studies involved tasks requiring subjects to compare static line drawings or photographs of objects with their associated sound. The ability to successfully perform this task depends entirely on the subjects' prior knowledge that the sight and sound of the object in question are related, and so is entirely dependent on semantic processes and memory. Therefore only regions involved in high-level cognitive stimulus evaluation could be isolated in such an experiment as opposed to the early perceptual crossmodal integration that was claimed. However, dynamic video footage was used in

part of the Beauchamp study, confirming the enhanced activation at the pSTS/MTG during bimodal versus unimodal stimulus processing (Beauchamp et al., 2004b).

2.6 Summary

This literature review has outlined the key theoretical concepts and empirical evidence that influenced the construction of the novel fMRI experiments described below. The explicit goal was to highlight the methodological pitfalls and assumptions made by some researchers and the solid foundations provided by others.

In the first fMRI experiment that comprises this thesis, the neural correlates of crossmodal integration were investigated by exposing subjects to video footage of commonly encountered stimuli in the environment and statistically contrasting regional brain responses under bimodal versus unimodal conditions. As the visual and auditory stimuli arose from the same physical event there was strict temporal covariation between dynamic changes in both stimuli. Therefore under bimodal conditions subjects perceived unified visual-auditory objects, whilst under unimodal conditions they perceived separate visual and auditory objects. Using an appropriate statistical technique to isolate brain areas exhibiting both a positive interaction between visual and auditory processing and an elevated responses to bimodal versus unimodal stimulus processing, it was possible to reveal the neural correlates of crossmodal integration. In section 2.3.2 the electrophysiological principles of crossmodal integration were outlined. An attempt was made to take advantage of the law of inverse effectiveness (Meredith and Stein, 1986) by degrading the quality of the video footage with visual and acoustic white noise in order to encourage proportionally larger bimodal than unimodal responses. In other words, visual and auditory stimuli that were minimally effective in allowing accurate stimulus identification under unimodal conditions were used, in order to exaggerate superadditive responses that might have resulted from crossmodal integration during bimodal conditions.

Another excellent example of how the chosen experimental paradigms were guided by previous experiments was to avoid the conflicting approaches taken by the Bushara and Calvert groups (Bushara et al., 2001; Calvert et al., 2001) when defining the neural correlates of crossmodal synchrony and asynchrony. The Bushara paper identified brain areas more active during crossmodal asynchrony than synchrony. The Calvert study on the other hand, in following the electrophysiologically defined

principles of crossmodal integration to the letter, made specific assumptions in their analytical approach that precluded the detection of such areas. This analysis was only able to identify areas at which greater activity was observed during crossmodal synchrony than under unimodal conditions (superadditivity) and where less activity was observed during crossmodal asynchrony (response depression). The aim was to settle these issues in experiment 2, by performing an fMRI study investigating the neural correlates of crossmodal synchronicity using a method that allowed both the questions asked by the Bushara study (where is there more activity during crossmodal asynchrony than synchrony?) and the Calvert study (where is there more activity during bimodal stimulation than unimodal stimulation?) to be confronted in a single experiment. This question was expanded in experiment 3 by investigating, for the first time, whether brain areas sensitive to synchronicity *between* the senses (crossmodal) were also sensitive to synchronicity *within* individual sensory modalities (intramodal).

In the fourth and last fMRI experiment, the neural correlates of visual and auditory beauty evaluation were investigated in order to pursue the development of a new branch of neuroscience research – neuroaesthetics - inaugurated by a recent study from this laboratory (Kawabata and Zeki, 2004). This enabled the characterisation of supramodal brain areas involved in the evaluation of both visual and acoustic beauty.

Part 3:

METHODS

&

RESULTS

3.1 Methodological overview

Four functional Magnetic Resonance Imaging studies comprise the empirical research upon which this thesis is based. The first subsection of this methodological overview (3.1.1) gives a basic description of the different stimuli used in each of the experiments and the tools used to construct them. The following section (3.1.2) provides technical information about the two MRI scanners used in these experiments, along with the arrangements for stimulus delivery and subject response acquisition during scanning. The final section in this overview (3.1.3) details the various aspects of MRI data processing, from preparing the raw scanner output for entry into Statistical Parametric Mapping (SPM) analyses, to the statistical techniques employed to contrast brain activations during different experimental manipulations. Care has been taken to delineate the main commonalities and differences in the approach to each experiment.

3.1.1 Stimulus development

The stimuli for all experiments were developed using an RM Innovator PC (PC1000-256), running the Windows 2000 operating system with a 1 GHz Pentium III processor, 256 MB RAM and a 50 GB Hard Disc (Research Machines, UK). Visual stimuli were displayed on a Sony Trinitron CPD-G520 monitor running at 85Hz and sound stimuli were presented via headphones (Vestax, UK) connected to the PC's audio output port.

In experiment 1, the visual-auditory object recognition study, the Premiere video editing software package (version 6.5, Adobe, US) was used to manipulate movie footage of everyday occurrences in urban settings, from which 5s visual and auditory clips of varying degrees of salience were created. The visual and auditory stimuli for experiments 2 and the visual stimuli for experiment 3, the visual-auditory synchronicity studies, were created entirely using software tools developed in-house by the Vislab computer programmer John Romaya (COGENT 2000, www.vislab.ucl.ac.uk) running in Matlab (Mathworks Inc., US). The auditory stimuli of experiment 3 were created using a synthesiser / keyboard (ROLAND Phantom XA, UK). In experiment 4, the visual-auditory neuroaesthetics study, a selection of visual artworks were manipulated in Premiere and Photoshop (version 6.0, Adobe, US), whilst music clips were digitised using ACE software (freeware available from: www.mp3-ripper.com) and edited in the Audacity software package (freeware available from: www.sourceforge.org).

During scanning, all experiments were controlled by bespoke Matlab scripts using COGENT commands. These programs randomised the stimulus order, delivered stimuli at the appropriate intervals, logged subject responses, stimulus onsets/offsets and experimental parameters and saved a Matlab file (MAT) at the end of each scanning session containing all experimental data for each subject.

Experiment 1: stimuli for the visual-auditory object recognition study

In experiment 1, subjects were asked to identify noise-degraded video footage of everyday environmental events from a choice of four possible options. The aim was to use stimuli that were minimally effective in permitting identification under unimodal conditions, in order that bimodal responses might induce enhanced regional brain activations, as per the law of inverse effectiveness (see section 2.3.2). To degrade the salience of visual and auditory clips, visual and acoustic white noise was mixed in various proportions to the raw video footage and sound tracks, respectively. Randomly generated visual white noise was created on a frame-by-frame basis in MATLAB to produce bitmap files (BMP), which were automatically loaded into Premiere as an animation and combined with the video footage using the 'opacity' rubberband. Auditory white noise was also created in MATLAB and was combined with the soundtrack of each clip using the volume rubberband. In this way the salience of the moving images and sound tracks could be manipulated at will. During scanning, subjects were presented with three different versions of twenty different video clips and of twenty different soundtracks, under unimodal and bimodal conditions. The aim was to find the neural correlates of visual-auditory integration of dynamic events in the natural environment by assessing which areas were more active during bimodal versus unimodal stimulation and also demonstrated a crossmodal factorial interaction. A baseline condition consisting of pure visual and auditory white noise was also included to enable this crossmodal interaction analysis. These stimuli are described in greater detail in the experiment 1 methods section (section 3.2).

Experiment 2: stimuli for the visual-auditory synchronicity study

In experiment 2, subjects were required to distinguish between synchronous and asynchronous, abstract, visual-auditory stimuli, which cycled continuously between two extremes, through all intervening values. The visual stimuli in this experiment consisted of a centrally presented, square, random dot array in which either dot colour, or dot

velocity, was varied continuously between two extremes to produce repeating cycles. Colour cycles consisted of repeating, gradual changes in dot colour from cyan to magenta and back to cyan again, through all intermediate colours. Motion cycles consisted of repeated progressions in unidirectional movement of the random dot array from low to high velocity and back to low velocity, once again through all intermediate values. Auditory stimuli consisted of a single pure tone varying in pitch between a high and low frequency, continuously through all intervening frequencies.

Visual colour with auditory pitch cycles, or visual motion with auditory pitch cycles, comprised the two types of bimodal stimuli. In the scanner, subjects were exposed to synchronous and asynchronous versions of these two types of bimodal stimuli. In addition to bimodal conditions subjects were also exposed to the same stimuli, but with the auditory component silenced (unimodal visual epochs), with the visual component rendered static and colourless (unimodal auditory epochs), or in the absence of stimulation in both modalities (rest epochs). These stimuli are described in greater detail in the experiment 2 methods section (section 3.4).

Experiment 3: stimuli for the crossmodal versus intramodal synchronicity study

In experiment 3, the stimulus array consisted of two visual stimuli presented concurrently with two auditory stimuli in every epoch. Subjects were asked to assess whether this stimulus array was intra-modally synchronous, inter-modally (or as it is described in this thesis – crossmodally) synchronous or completely asynchronous. The difference between each condition was established purely on the basis of the degree to which the cyclical changes between each of these four sensory channels were temporally correlated (synchronous) or decorrelated (asynchronous). The advantage of this set up over the previous experiment is that in addition to crossmodal synchrony and asynchrony, where visual and auditory cycles were temporally aligned or desynchronised, it was also possible to control intra-modal timing relationships. Intra-modal synchrony describes simultaneous changes between the two visual stimuli and between the two auditory stimuli. Synchrony could therefore be introduced intramodally between the two visual stimuli and between the two auditory stimuli, or crossmodally between visual and auditory stimuli, on an independent basis. These stimuli are described in greater detail in the experiment 3 methods section (section 3.6).

Experiment 4: stimuli for the visual-auditory neuroaesthetics study

In experiment 4, subjects evaluated the aesthetic beauty of a selection of visual art and music extracts on a scale of 1-8. The aim was to investigate the supramodal neural correlates of beauty common to both the evaluation of visual and acoustic art forms, by identifying brain areas at which activity is parametrically modulated as a function of perceived beauty. By its very nature music gradually unfolds over time and so it was necessary to select clips of several seconds duration. This ensured that subjects would give genuine rather than arbitrary aesthetic ratings for each piece. Twelve-second clips were edited from forty different tracks of classical/jazz music and comprised the acoustic stimuli for this experiment. For equivalence, the visual stimuli were presented for the same duration. However, instead of presenting the same static image for an extended period, Premiere was used to create an animation of each artwork, which produced the illusion that the screen was gradually zooming in to the central area of the image, throughout the 12s duration. These stimuli and the experimental paradigm are described in greater detail in the experiment 4 methods section (section 3.8).

3.1.2 Magnetic Resonance Imaging (MRI) scanning

Subject recruitment

All subjects were recruited by means of advertising campaigns consisting of posters at cash machines, on notice boards and in eateries of the University of London Union, Birkbeck College and University College London. Subjects were invited to participate in the scanning experiments if they had no prior history of neurological injury or psychiatric illness, normal hearing and normal or corrected to normal vision by self-report. All subjects gave written informed consent for the experimental procedures, which were approved by the National Hospital for Neurology and Neurosurgery Ethical Review committee, Queen Square, London.

Stimulus delivery and subject responses

Visual stimuli were back projected from an LCD projector (NEC LT-158, US) of brightness 1500 lumen per square metre onto a screen mounted behind the subject's head. Subjects assumed a supine position in the scanner bore and viewed the projector screen via an angled mirror mounted on the head coil, which was positioned approximately (~) 16mm from the eye. The distance from mirror to the screen was ~614mm so that a visual array of 85 x 85mm subtended ~8° of visual angle and a 2mm dot projected onto the screen subtended ~0.2°. During scanning subjects were played the acoustic stimuli binaurally at 90dB SPL over the ambient scanner noise. Custom-built electrostatic headphones, installed into industrial standard ear defenders (3M, US) with noise reduction rating of 23dB, were used to deliver the acoustic stimuli in experiment 1 (which required superior sound quality). The pneumatic headphones provided with the scanner (Siemens, Germany) were used during experiments 2, 3 and 4. Subjects also wore earplugs with the headphones as this was found to further attenuate the ambient scanner noise without causing noticeable deterioration of the acoustic stimuli, thus further emphasising the experimental sound stimuli.

In all experiments subjects were required to make discriminations of some description such as: stimulus identity (Experiment 1), stimulus synchronicity (Experiments 2 & 3) or aesthetic rating (Experiment 4). In experiments 1, 2 and 3, subjects were provided with a right-handed, four-button response box only, in order that they could make their response choices known during a 2s post-stimulus button-press epoch using buttons 1-4 (experiment 1), 1-2 (experiment 2) or 1-3 (experiment 3)

according to 4-alternative forced choice (AFC), 2-AFC or 3-AFC protocol, respectively. In experiment 4, subjects used both a left- and a right-handed, four-button, response box, enabling them to select from eight possible keys to register their chosen aesthetic rating (from 1-8). For the first three experiments using the right-handed, button box only, button one was pressed with the forefinger, button two with the middle finger, button three with the ring finger and button four with the little finger. In experiment 4, buttons 1-8 ran consecutively from left to right with the button furthest left (pressed with the left hand little finger) as button 1 and the button furthest right (pressed with the right hand little finger) as button 8.

Acoustic scanner emissions

As standard echo planar imaging (EPI) sequences involve fast gradient switching, a continuous high-pitched bleeping sound is emitted, so many fMRI studies investigating regional brain responses to auditory stimuli use sparse sampling methods. Sparse sampling EPI sequences take advantage of the ~6s lag of the blood oxygen level dependent (BOLD) response behind the neuronal activity that induces it. Such periodic scanning protocols alternate between silent periods – when brief auditory stimuli are presented, followed by subsequent volume acquisitions - during which the acoustic noise pollution is produced. For several reasons the standard, continuous EPI sequence was used instead in all four experiments, with all sound stimuli presented concurrently with the high-pitched bleeping sounds produced by the scanners. Firstly, the primary focus of all studies was the interaction between auditory and visual stimuli rather than specific responses to the auditory stimuli themselves. Secondly, the auditory stimuli in experiments 2 and 4 were relatively long lasting (12s duration) making them unsuitable for sparse sampling methods. The considerable lengthening of scan time necessitated by the use of sparse sampling paradigms seemed unwarranted in experiments 1 and 3, as subjects reported no difficulty in hearing the auditory stimuli over the sound of the continuous EPI sequence during pilot studies and earlier fMRI experiments.

MR Scanning information

All four experiments were performed using one of two different scanners, both located at the Functional Imaging Laboratory (Wellcome Department of Cognitive Neuroscience, 12 Queen Square, London WC1N 3BG: www.fil.ion.ucl.ac.uk). Experiments 1, 2 and 3 were conducted on the Magnetom Allegra 3 Tesla scanner

(Siemens, Erlangen, Germany), whilst experiment 4 was conducted on the Magnetom Sonata 1.5 Tesla scanner (Siemens, Erlangen, Germany).

Functional Magnetic Resonance Imaging (fMRI) using the BOLD contrast has made possible the rapid, non-invasive and high spatial resolution imaging of the human brain. In all experiments, a T2*-weighted, gradient echo-planar imaging (EPI) sequence was used to maximise the BOLD contrast in whole brain acquisitions consisting of descending transverse slices of 3 x 3 x 2mm voxels. The EPI sequence on the 1.5T Sonata scanner acquired a single slice every 90ms (TE=50ms), whilst the 3T Allegra scanner acquired a single slice every 65ms (TE=30ms) and on both, the interslice distance factor was always 50%. The EPI sequences on both scanners used a flip angle of 90°, anterior to posterior phase encode direction and had a 192mm field of view (FoV read) comprising 64 x 64 matrix. The number of slices and slice thickness was adjusted according to the epoch length during each experiment to maintain full brain acquisition without subsampling the BOLD response for each stimulus.

Functional scans were always divided into two sessions to reduce the length of any one sitting so that the subjects would maintain full concentration throughout. These two sessions were separated by a T1-weighted structural image acquisition, which provide high-resolution (1 x 1 x 1mm) anatomical detail with which functional activations could be co-registered and overlaid. This enables accurate visualisation of topographical position of the activations, when required. On the Sonata scanner the structural sequence consisted of 160 x 1mm saggital slices, each acquired every 16ms (TE=9ms), with a flip angle of 25° and a 256mm FoV. On the Allegra scanner the structural sequence consisted of 176 x 1mm saggital slices, each acquired every 7.92ms (TE=2.4ms), with a flip angle of 15° and a 256mm FoV.

3.1.3 Data Analysis

Data analysis I: pre-processing

All MRI data was processed entirely within the SPM2 software package (Wellcome Department of Cognitive Neuroscience, 12 Queen Square, London WC1N 3BG: www.fil.ion.ucl.ac.uk). In all studies, pre-processing essentially consisted of trajectory based reconstruction, spatial (and in experiment 2 also temporal) realignment, normalisation to the standard SPM2 EPI template and spatial smoothing with a 12mm full-width at half-maximum (FWHM) 3-D Gaussian kernel. T1-weighted structural images were co-registered with the mean EPI image on a subject-by-subject basis and then spatially normalised according to each subject's individual deformation parameters recorded from the functional normalisation procedure, in order for the anatomical position of overlaid activations to be accurately portrayed.

In order to combine functional data from different scans from the same subject, or data from different subjects, they must conform to the same anatomical frame of reference. Thus the sequence of spatial transformations and morphological operations required to establish this generic anatomical frame of reference are described here. Data were inspected visually to ensure that the trajectory-based reconstruction process had been successful. The first few scans (referred to here as 'dummy scans') were discarded to allow for T1 equilibration effects. Spatial realignment involved estimation of the six parameters of an affine "rigid-body" transformation (x, y, z transformations and roll, yaw and pitch rotations) that minimises the sum-of-squared differences between each successive scan and a reference scan. Images were thus realigned to the first post-dummy scan by applying these translation and rotation parameters using sinc interpolation, in order to correct for head movement during acquisition. Each of these transformation parameters was saved and used later as measures of subjects' head movements for use in the design matrix to further compensate for residual movement effects. These realignment parameters were also visually inspected for excessive (shifts of over 2 voxel widths i.e. < 6mm) or periodic movements that might have correlated with different stimulus epochs. Had any excessive or periodic movements of this nature occurred, subject data would have been rejected. This was not necessary in any of the experiments presented here as subject movements stayed within the proscribed limits.

After realignment, the data were again visually inspected to ensure that the process had been successful.

As different slices comprising the complete brain volume were acquired sequentially every 65ms (Sonata 1.5T scanner) or 90ms (Allegra 3T scanner) over a period of 2.08-4.32s (depending on the number of slices / slice thickness parameters used to capture each single full brain volume), the realigned data could then be time sliced, that is, temporally realigned to the middle slice, using sinc interpolation over time to ensure that the data from any given volume were sampled at the same time.

After spatial and temporal realignment, the mean spatially realigned image of the scan time series was used to estimate the linear and non-linear warping parameters that would transform the functional images to the standard EPI template provided in SPM2. This is similar to the average of 305 brains created by the Montreal Neurological Institute (MNI), which approximates to the co-ordinate system used by the Talairach brain atlas (Talairach and Tournoux, 1988). Bilinear interpolation was used to warp each image according to the estimated linear and non-linear warping parameters and resliced to 2 x 2 x 2mm voxel size. This convention has been established in SPM analyses in order to enable co-ordinates of a particular voxel to be reported in such a way that allows comparison to those reported in previous studies that have also used this spatial co-ordinate system.

In the text of this thesis, specific neuroanatomical co-ordinates are always enclosed in square brackets (e.g. the origin is at co-ordinate [0, 0, 0, MNI]) where the sequence of 3 digits denotes the distance in millimetres from the origin along the x, y and z axis, followed by the initials 'MNI' to indicate the stereotaxic framework in which the coordinates should be applied.

The next step was to smooth the data, which may seem slightly counterintuitive in terms of degrading spatial resolution for which fMRI is lauded, but is in fact essential in fMRI analyses for several reasons. Firstly, by the central limit theorem, smoothing the data will render the errors more normal in their distribution and ensure the validity of inferences based on parametric tests (as Normal distribution of errors is assumed in making such inferences). Secondly, as Gaussian random field theory is employed to calculate the number of independent resolution elements (RESELS) for use in a Bonferroni-like correction for multiple comparisons, which assumes that the error terms are a reasonable lattice representation of an underlying and smooth Gaussian field, it is required that smoothness be substantially greater than voxel size (Friston et al., 1991).

Thirdly, as inter-subject averaging is used in the generation of the results reported here and cerebral anatomy varies from subject to subject, it is often necessary to smooth even more, to project the data onto a spatial scale where homologies in functional anatomy are expressed among subjects. Hence, data were spatially smoothed with a Gaussian kernel of 12mm FWHM to satisfy these statistical assumptions and allow for inter-subject anatomical variation.

For every slice a global voxel value was calculated and removed from the data, a correction that removes some of the slowly varying scanner noise. These pre-processed multiple voxel time-series were then convolved with the canonical haemodynamic response function (HRF) and band-pass filtered with a low-frequency cut-off of 128s and a high-frequency cut-off shaped to the characteristics of the HRF. This filtering allowed removal of much of the remaining slowly varying scanner noise, as well as high frequency effects such as cardio-vascular coupling or breathing-related noise.

Data analysis II: statistical analysis in SPM2

Statistical parametric mapping (SPM) is a mass univariate approach that calculates a statistic for every single voxel in the brain using the general linear model (GLM). The experimental manipulations were specified in a model or a design matrix, which was then fitted to every voxel comprising the whole brain volume. This enables the size of the experimental effect to be estimated providing the basis for specific hypotheses or contrasts to be tested and statistical inferences reached.

Separate boxcar functions, indicating the onset and offset times for each different category of experimental manipulation, were convolved with the HRF. The six realignment parameters obtained during the realignment pre-processing stage were also modelled as nuisance variables, but were not convolved with the HRF and were simply entered into the multiple regression as one value per scan. All regressors were then entered into a multiple linear regression analysis with the BOLD signal from each voxel from the image in turn (mass-univariate approach). Each regressor was fitted to the GLM (Equation 1) using a least-squares estimation approach (Friston et al., 1995):

$$y = \mu + \beta.X + \varepsilon \qquad \textbf{Equation 1.}$$

Where: y = scan data

μ = intercept constant (i.e. mean activity of that voxel over scanning session)

β = the parameter estimate (i.e. the slope of the regression)

X = linear combination of explanatory variables (i.e. the basis functions modelling the experimental paradigm)

ε = independently and identically (Normally) distributed residual error

The parameter estimates were calculated from the ordinary least squares fit to the data for each regressor, explaining a certain proportion of the variance in the BOLD signal. The model also fitted the intercept constant. Statistics could then be generated under the null hypothesis that the regression slope (β) was 0, or that the difference between regression slopes for different basis functions was zero (β 1- β 2 = 0). Indeed by setting an appropriate vector of contrast weights, much more complex comparisons were possible (e.g. [β 1- β 2] - [β 3- β 4] = 0 etc). The difference between actual regression slopes could then be compared to the residual error at each voxel to generate a t-statistic according to equation 2:

$$t = \beta 1 - \beta 2 \,/\, \mathrm{sqrt}\left[s^2 . \left(1 / n_1 + 1 / n_2 \right) \right] \qquad \textbf{Equation 2.}$$

Where β 1and β 2 are the regression slopes to be contrasted, s^2 is the variance of the residual error and n_1 and n_2 are the number of observations (scans) used to construct each β. The degrees of freedom (df) were established according to equation 3 to enable the t-statistics to be converted to a probability value using Student's t-distribution.

$$df = \left[(n_1 - 1) + (n_2 - 1) \right] / 2 \qquad \textbf{Equation 3.}$$

Using these equations at every voxel, SPM{t} maps were produced covering the entire brain. This map was then thresholded at an appropriate p-value to reveal voxels of interest.

Data analysis III: fixed effects or random effects analyses?

A subject's BOLD response varies from trial to trial and this response also varies from subject to subject. Both within-subject and between-subject variability must be taken into account when making inferences about the population. There are two

different statistical approaches that enable the effects measured in the individual to be generalised across different subjects: fixed effects (FFX) and random effects (RFX) analyses (Penny & Holmes, 2004).

A fixed effects analysis was used in experiment 2 only. The data for all subjects were concatenated and entered into the GLM as a single column vector. A multisubject design matrix involving three regressors per condition comprising the boxcar function convolved with 1) the canonical HRF, 2) the temporal derivative and 3) the dispersion derivative, was then fitted to the data. After model estimation, as all subjects were modelled together, the desired contrasts were simply set up at the first level incorporating parameter estimates from all subjects at the same time. In order for a linear sum of the three regressors to be accounted for in the statistical analyses, SPM{F} maps were created. The F-statistic highlights voxels with significantly non-zero differential activity, reflecting the 'average effect in the group'. This means that when contrasts were performed between two stimulus conditions (say A>B), in the context of a FFX analysis, a single very large parameter estimate for condition A in one subject could have skewed the average group effect. This leads to the possibility of rejecting the null hypothesis even though the effect was not demonstrated in most subjects. This problem arises because fixed effects only take into consideration the within subject variability, not the between subject variability. However, by plotting each subject's parameter estimates, it could be established whether or not the majority of subjects contributed to the effect and so single subject biases could be ruled out, thus ensuring the validity of inferences made regarding significant voxels. The advantage of the FFX analysis is that it can be used in experiments in which data was collected from very few subjects, harnessing the large degrees of freedom and low scan-to-scan variability of first level SPM analyses when assessing the significance of an estimated response. The drawback is that, as the between-subject variability is not taken into in account with this method, it is not possible to make formal inferences about the population from which the subjects were drawn unless conjunction analyses are performed (Friston et al., 1999).

The random effects in a RFX analysis are the subjects, as they are randomly drawn from the 'local' population, allowing the sampling variability to be taken into account and so inferences can be made about the population from which the subjects were drawn. RFX analyses were used in experiments 1, 3 and 4, where the data for each subject was entered into separate GLMs at the first level, each with identical design

matrices and, after model estimation, contrasts between conditions of interest were performed yielding subject-specific contrast images (CON). A single CON image per subject was then fed into a second level GLM that produced a 'summary statistic' by means of a one-sample t-test, which established at which voxels significantly different activations occurred across all subjects. The advantage of the RFX approach is that the resulting summary statistic pertains to a null hypothesis that there is no population effect at that voxel, and so rejection of the null hypothesis allowed us to infer that the significant differential activity would be observed in a significant proportion of the population from which the subjects were drawn. This effectively removed the problem of individual subject bias inherent in FFX analyses and eliminated the need to check consistency across subjects by plotting individual parameter estimates. The findings can also be generalised beyond the subject sample to the population at large. The disadvantage of this approach was that at the second level, the degrees of freedom were calculated from the number of subjects, rather than the number of scans (in the case of first level analyses) leading to: a) highly conservative statistics, and b) the need for at least twelve subjects for significant results.

In order to perform conjunction analyses using a one-way ANOVA, it was necessary to take more than one contrast per subject up to the second level and so one must take into account that the contrasts may be correlated, or of unequal variance. This entails estimating the departure from the assumption of identical and independent distribution of the residual errors, i.e. the 'non-sphericity', and making an adjustment to the degrees of freedom used to calculate the statistic accordingly.

In experiment 2, where a FFX analysis was used, if there were no *a priori* hypotheses regarding which neuroanatomical areas were expected to be differentially activated, results were statistically thresholded at $p<0.05$ with a correction for multiple comparisons. If an anatomical *a priori* hypothesis existed regarding where activity might reasonably be expected, or when using conservative statistical techniques such as conjunction analyses, an uncorrected statistical threshold of $p < 0.001$ was used. In experiments 1, 3 and 4, where the random effects analyses were used, an uncorrected statistical threshold of $p < 0.001$ was used, unless otherwise stated.

Data analysis IV: the crossmodal integration conjunction

Previous fMRI studies investigating the neural correlates of multisensory integration (e.g. Beauchamp et al., 2004b), or as it is described in this thesis: crossmodal

integration (CMI), have invariably used the crossmodal elevation contrast (VA>V+A). This isolates brain areas at which significantly greater responses are detected under bimodal (VA), as compared to unimodal (V+A), stimulus conditions. A more robust approach to the identification of putative sites of CMI invokes a two-by-two factorial design in which visual and auditory sensory stimulation are the factors and the presence or absence of the stimulation are the levels of these factors (figure 5). The factorial interaction between visual and auditory stimulation is the most efficient statistical approach for detecting brain areas involved in crossmodal integration (Friston, personal communication).

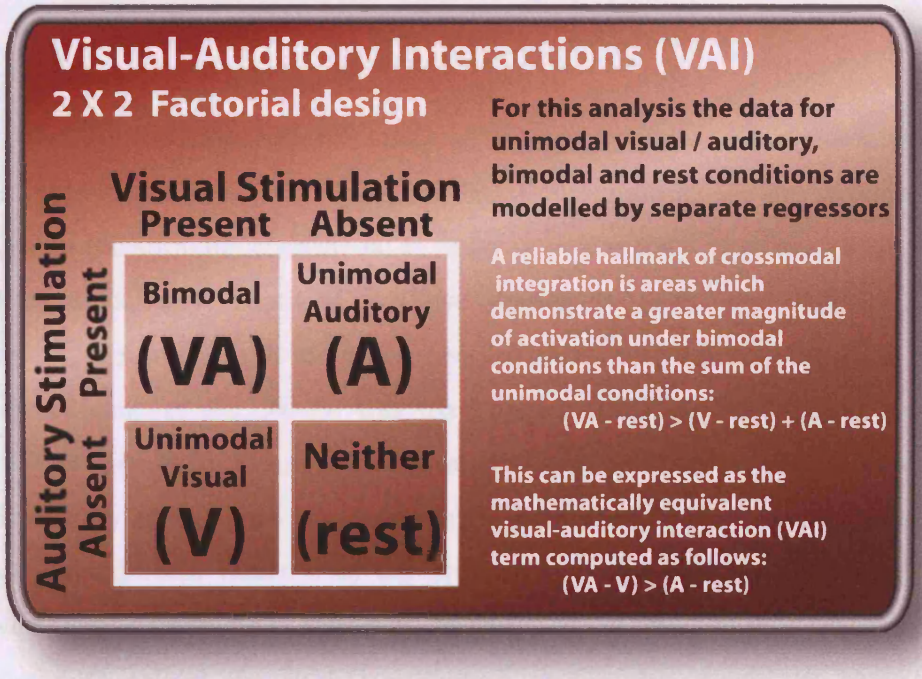


Figure 5: Factorial design for detecting visual-auditory interactions. This is a powerful statistical approach enabling the identification of brain areas at which visual stimulus processing is significantly modulated by auditory stimulus processing and *vice versa*.

Voxels at which significant visual-auditory interactions (VAI) are detected reflect activity profiles where responses to auditory stimulation are significantly modulated by the presence of concurrent visual stimulation (and *vice versa*). The interaction contrast in this context tests the null hypothesis in equation 4:

$$H_o: \quad (\beta_{VA} - \beta_V) - (\beta_A - \beta_{REST}) = 0 \qquad \textbf{Equation 4.}$$

Where: β_{VA} = parameter estimate for concurrently presented bimodal stimuli

β_V = parameter estimate for visual stimulation only

β_A = parameter estimate for auditory stimulation only

β_{REST} = parameter estimate for neutral visual and auditory stimulation

This is equivalent to a statistical test for voxels significantly more active during bimodal stimulation than unimodal stimulation, having normalised their parameter estimates to the baseline, or rest condition. This can be achieved by testing the mathematically identical null hypothesis in equation 5:

$$H_o: (\beta_{VA} - \beta_{REST}) - [(\beta_V - \beta_{REST}) + (\beta_A - \beta_{REST})] = 0 \qquad \textbf{Equation 5.}$$

Despite the statistical efficiency of the VAI approach it is theoretically possible that it could reveal voxels in which the baseline condition elicited stronger responses than any of the stimulus conditions. In order to rule out the possibility that such voxels might be mistakenly identified in these studies as neural correlates of crossmodal integration, the crossmodal elevation contrast (VA > V+A) used by other researchers (e.g. Beauchamp et al., 2004b) was used as a control. By performing a conjunction analysis between the VAI and crossmodal elevation contrasts, we could be sure that the visual-auditory interactions were driven by increased activity during bimodal versus unimodal conditions, rather than the artifactual causes described above. Voxels that survived this conjunction analysis, referred to as the 'CMI conjunction', were considered to reflect the BOLD activation profile that would be expected during crossmodal integration.

Response depression, that is reduced bimodal activity versus unimodal activity during incongruent bimodal stimulation, is also used as an indicator of CMI in animal electrophysiology, but is a less robust indicator in fMRI. One reason for this is that combined visual and auditory stimulation of separate neuronal populations within a voxel could potentially induce a saturation effect in the detected BOLD response. This would result in a potentially high incidence of false positives, creating the inaccurate impression of bimodal response depression when contrasted against the sum of the unsaturated unimodal responses. The BOLD 'ceiling effect' may of course also adversely affect the chances of detecting positive visual-auditory interactions, but use of

such a conservative approach should serve only to bolster support for inferences of CMI in these experiments.

Data analysis V: unimodal, crossmodal and bimodal contrasts

Where possible, a standard approach has been employed to allow differentiation of brain areas involved in the processing of unimodal as opposed to bimodal sensory inputs. Firstly, in experiments 1, 2 & 4, unimodal visual (V) and unimodal auditory (A) epochs were included and so simple subtraction analyses could be used to identify the neuroanatomical territories involved in unimodal visual versus unimodal auditory sensory perception. Auditory areas demonstrated greater BOLD responses during unimodal auditory than visual stimulation (A>V), whilst visual areas were identified by voxels that were more active during unimodal visual than auditory stimulation (V>A). These unimodal contrasts were performed, not with the intention of finding new unimodal brain areas, but rather so that areas exhibiting response profiles indicative of crossmodal integration would be accurately identified as residing within or beyond the territory of unimodal cortex. Thus, having first characterised regions in which unimodal processing is accomplished, the crossmodal and bimodal contrasts were then considered. The CMI conjunction was then performed, which, as described in the previous subsection (data analysis IV), reveals voxels that survive both the VAI contrast ([VA-V] > [A-Rest]) and the crossmodal elevation contrast (VA>V+A), to identify brain areas exhibiting positive interactions between visual and auditory sensory processing. In experiments 2, 3 and 4 other contrasts were performed according to the specific aims of each particular study i.e. synchrony/asynchrony detection (experiments 2 and 3) and parametric beauty modulations (experiment 4).

3.2 Experiment 1 methods: visual-auditory object recognition

3.2.1 Experiment 1 aims

This study was designed explicitly to reveal regional brain activations resulting from crossmodal integration of dynamic sensory information for the recognition of everyday environmental events. The visual and auditory parts of a selection of video clips were independently degraded using white noise in order to control object recognition performance. By using stimuli that were minimally effective in enabling object recognition under unimodal conditions, the aim was to apply the law of inverse effectiveness (Meredith and Stein, 1986) in an attempt to increase the relative magnitude of the BOLD response during bimodal versus unimodal conditions, thus revealing brain areas involved in crossmodal perceptual integration.

3.2.2 Experiment 1 subjects

Twelve subjects, between 18-29 years (mean 24.6 years SD ±2.9 years), seven of whom were male, were recruited according to standard criteria (section 3.1.2 - subject recruitment) for participation in the scanning experiment.

3.2.3 Experiment 1 stimuli

The video footage used in this experiment was recorded using a tripod-mounted DCR-TRV325E Digital Video Camera Recorder (SONY, Japan), captured to hard disk using the RTX.10 capture card (Matrox, Canada) and edited in Premiere 6.5 (Adobe, USA). Video footage was filmed around Central London, a local gym and University College London, mostly involving actors performing various tasks that create acoustic emissions (descriptions of which can be found in the appendix – section 5.1.1). The footage was then edited using Adobe Premiere to create a series of 5s clips depicting 20 different events. Visual and acoustic white noise was then proportionally mixed into the film footage, again using Adobe Premiere, in order to degrade the clarity of both visual and auditory parts making recognition more difficult (figure 6).

Multiple randomly generated visual white noise arrays of 320 x 280 pixels were created on a frame-by-frame basis using a simple MATLAB script (provided in

appendix 5.1.2) and the resulting bitmap (BMP) files were loaded into Premiere. The resulting visual white noise animation was proportionally mixed with each 5s clip of black and white rendered film footage, using the opacity rubberband in Premiere. The final clips were rendered as 320x280 pixel Microsoft multimedia format Audio Visual Interleaved (AVI) files of 25 frames per second created using square pixels. Acoustic white noise was created using a simple MATLAB function (preparewhitenoise.m), which was saved as a Microsoft audio format (WAV) file, loaded into Premiere and proportionally mixed with the audio tracks using the volume rubberband.

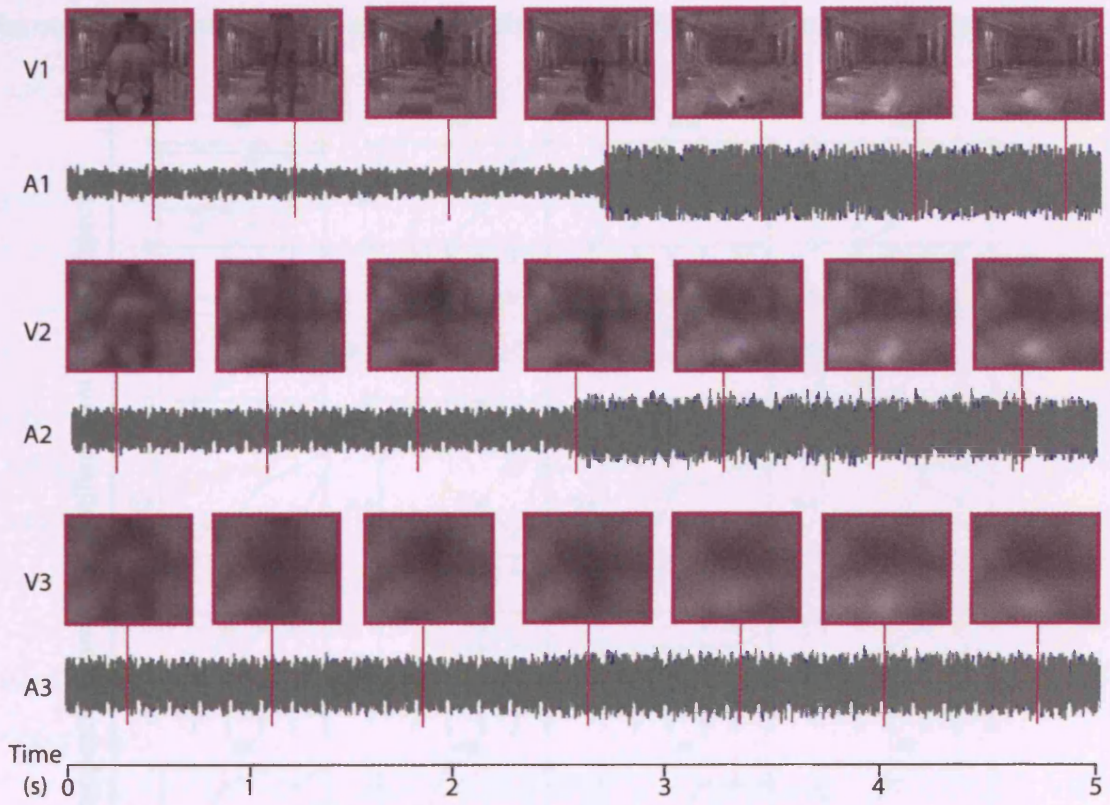


Figure 6: Mixing visual and auditory white noise into film footage. The visual (V) and auditory (A) movie footage of an actor jumping into a swimming pool, with no added noise (V1/A1), with a small amount of added noise (V2/A2) and with large amounts of added noise (V3/A3). Gaussian blur was used to create visual stimuli in this figure for illustrative purposes only.

A pre-scan psychophysical study was conducted, using a separate group of subjects (n=6, 3 male) from those that were to be scanned, in order to titrate the appropriate proportions of white noise that should be mixed into the visual and auditory stimuli to achieve equivalent levels of stimulus salience. A recording of the acoustic emission from the Allegra scanner (caused by the rapid switching of the gradient coils)

was also played over the headphones to simulate the experimental conditions of a scanning experiment for this pre-scan study. The psychophysical experimental procedure consisted of pseudo-randomly presented unimodal visual, unimodal auditory and bimodal visual-auditory stimuli, controlled by a MATLAB programme using COGENT commands. During the inter-stimulus interval, subjects were required to select, by means of a four-alternative forced choice (4-AFC) button press, which of four options listed on the monitor best described what they perceived. Mean across-subject performance was calculated for these subjects and the resulting psychophysical curves were used to titrate the amount of noise required to create unimodal stimuli that would be correctly identified with approximately 40%, 65% and 90% accuracy (figure 7).

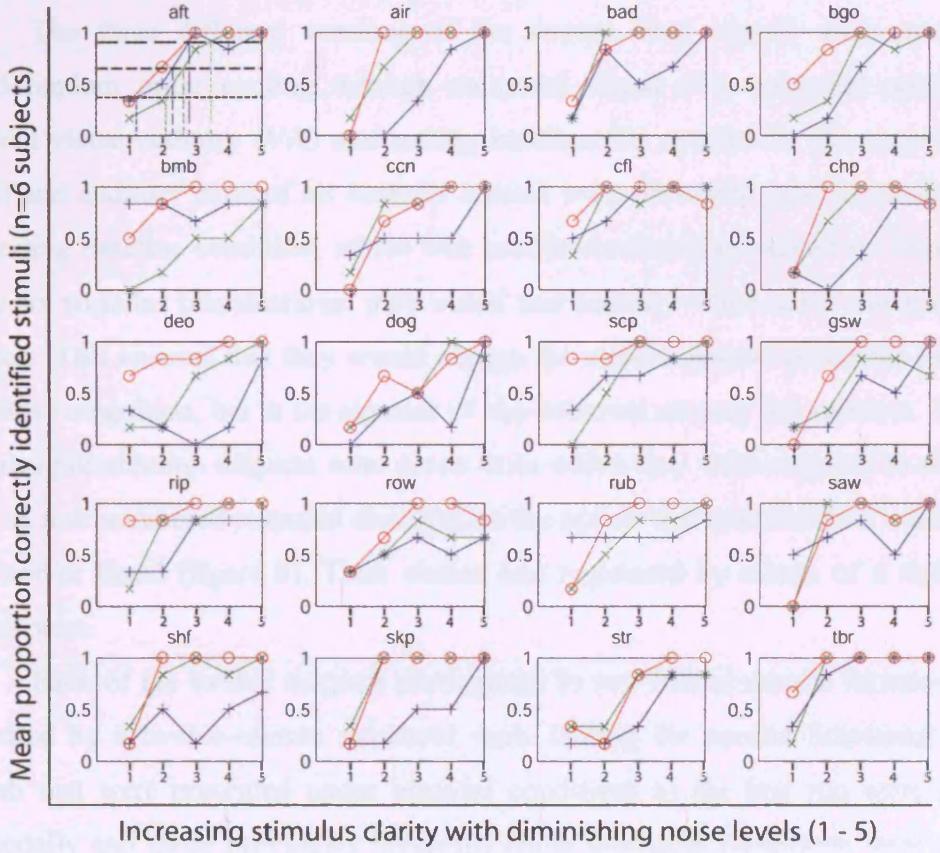


Figure 7: Pre-scan psychophysics. These twenty plots demonstrate the mean hit rate of six subjects required to identify twenty different film clips (plot ID codes described in section 5..1.1) across five levels of salience, under unimodal visual (green line & 'x' data point), unimodal auditory (blue line & '+' data point) and bimodal visual-auditory (red line & 'O' data point) conditions. The 'aft' stimulus data plot (top left) demonstrates how appropriate noise levels for the unimodal stimuli were calculated to produce scanning stimuli with standardised performance levels equivalent to 40% (small dashed lines), 65% (medium dashed lines) and 90% (large dashed lines) hit rates.

3.2.4 Experiment 1 scanning procedure

Functional scanning sessions consisted of 280 whole brain volumes, acquired using the standard T2*-weighted, echo planar imaging (EPI) sequence on the 3 Tesla Allegra MR scanner (Siemens, Erlangen, Germany). Each volume consisted of 48 slices, with a slice thickness of 2mm, giving a TR of 3120ms and included the entire cerebrum and cerebellum. All other scanning parameters have been described previously (section 3.1.2 – MR scanning information).

During scanning, visual film clips were presented on a grey background in a centrally presented rectangle of 6.8° height and 9.0° width (degrees of visual angle) and the acoustic stimuli were presented via electrostatic headphones at 90dB SPL (see section 3.1.2 – stimulus delivery, for further details).

The three different versions of the twenty final stimuli were presented in pseudorandom order cycling through unimodal visual (V), unimodal auditory (A), bimodal visual-auditory (VA) and resting baseline (R) epochs. In this experiment the visual and auditory parts of all bimodal stimuli were presented synchronously. During the resting baseline condition, which was pseudo-randomly presented on average once every six stimulus presentations, pure visual and auditory white noise was presented to subjects. This ensured that they would engage the same cognitive processes involved in stimulus recognition, but in the absence of any coherent sensory information. After each stimulus presentation subjects were given 4s in which they were required to select from a list of four numbered stimulus descriptions the option that best matched what they had seen and/or heard (figure 8). Their choice was registered by means of a right-handed button press.

Each of the twelve subjects participated in two fifteen-minute functional scans, separated by a twelve-minute structural scan. During the second functional scan, all stimuli that were presented under bimodal conditions in the first run were presented unimodally and those previously presented under unimodal conditions were presented bimodally. Whether stimuli were presented unimodally in the first run and then bimodally in the second, or bimodally in the first and then unimodally in the second run, was counterbalanced across subjects.

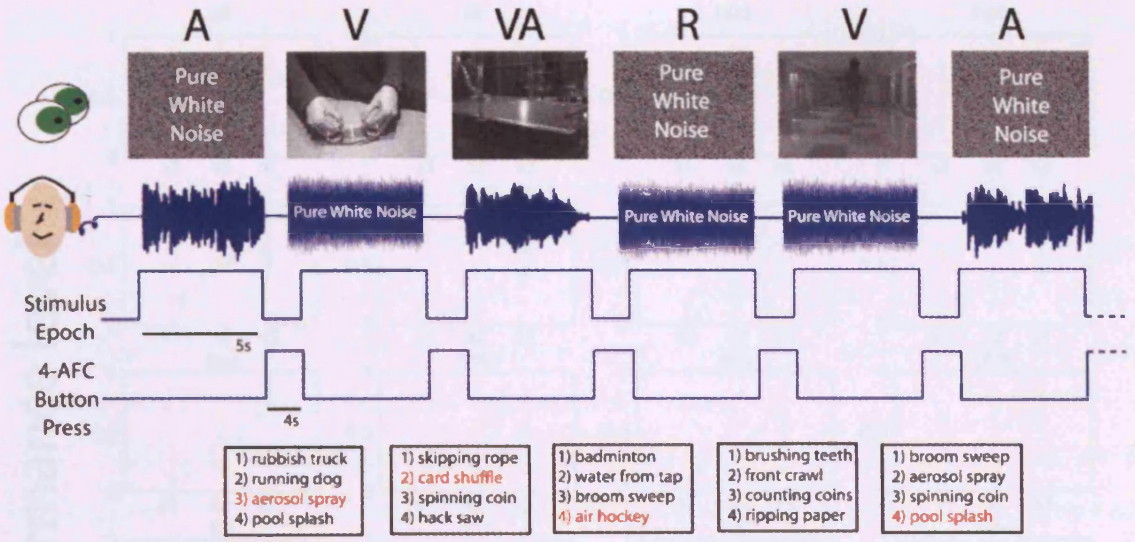


Figure 8: Experiment 1 stimulus presentation paradigm. Visual only (V), auditory only (A), combined visual and auditory (VA) and rest (R) epochs were presented to the subject in pseudorandom order. A 4s response period followed each stimulus during which subjects selected, via a four alternative forced choice button press, the option that best matched what had been perceived. For illustrative clarity, the correct answers have been highlighted in red and visual images / audio waveforms are portrayed without the added white noise. During scanning all stimuli were in fact embedded in white noise and the correct stimulus description was never indicated.

3.2.5 Experiment 1 data analysis

Mean across-subject performance was analysed to establish how well subjects were able to identify the various stimuli under unimodal visual, unimodal auditory and bimodal visual-auditory conditions (figure 9). It is clear from these psychometric curves that an unforeseen problem arose with many of the auditory stimuli, as subjects rarely identified many of the unimodal auditory stimuli with a success rate in excess of 50%, even at the highest level of salience, i.e. with the minimum amount of added noise (figure 9, blue dot-dashed lines). Although in a minority of cases the proportion of correctly identified auditory stimuli increased linearly with stimulus salience, as expected (e.g. the 'air', 'ccn', 'cfl', 'chp', 'deo' & 'skp' stimuli), several of the acoustic psychometric curves were completely flat (e.g. 'dog', 'row' & 'str'). In the few stimuli in which visual and auditory stimulus salience *was* well controlled, i.e. with well-aligned visual and auditory psychometric curves, the perceptual enhancement under bimodal versus unimodal conditions was robust (e.g. the 'bad' stimulus).

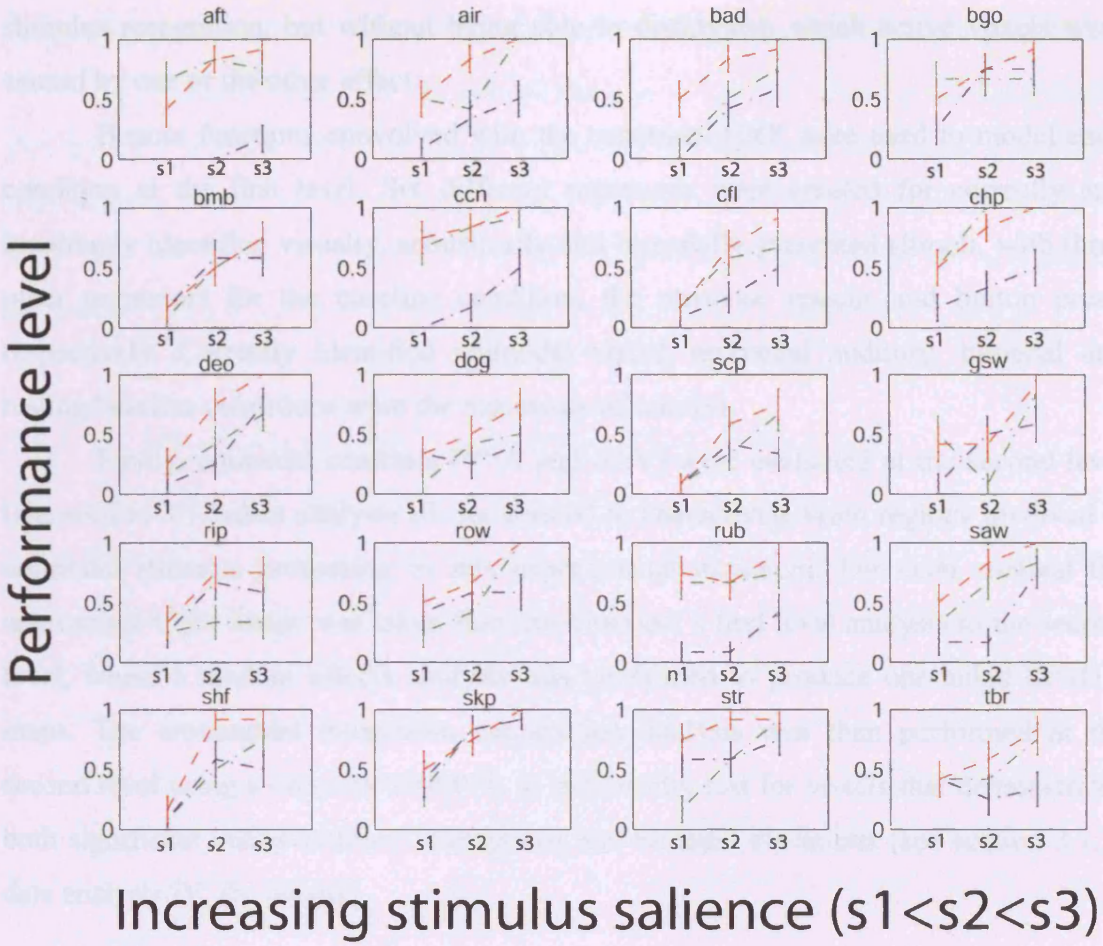


Figure 9: Mean recognition performance across twelve scanned subjects. Mean performance under unimodal visual (green, dot-dash), unimodal auditory (blue, dot-dash) and bimodal (red, dashed) presentation conditions for each of the twenty different stimuli under three levels of stimulus salience (s1 –high noise; s2 – medium noise; s3 – low noise).

Despite failing to produce the optimal psychophysical results that would have enabled a parametric analysis of visual-auditory integration profiles across the three different levels of stimulus salience, it was still possible to perform a valid default analysis. This was made possible by modelling each subject's data according to their own personal performance profile in a random effects analysis. Correctly and incorrectly identified unimodal visual, unimodal auditory and bimodal stimuli were all modelled separately. This enabled potential confounding factors to be avoided such as those that might have arisen had statistical contrasts been made without taking account of recognition accuracy. For instance the visual versus auditory contrast would have revealed both regional brain activations induced by visual stimulus processing and

stimulus recognition, but without being able to distinguish which active voxels were caused by one or the other effect.

Boxcar functions convolved with the canonical HRF were used to model each condition at the first level. Six different regressors were created for correctly and incorrectly identified visually, acoustically and bimodally presented stimuli, with three other regressors for the baseline condition, the response epochs and button press, respectively. Correctly identified unimodal visual, unimodal auditory, bimodal and resting baseline conditions were the regressors of interest.

Firstly, unimodal contrasts (V>A and A>V) were evaluated at the second level (see section 3.1.3-data analysis III, for details) to characterise brain regions involved in unimodal stimulus processing in this experimental paradigm. For each contrast the appropriate CON image was taken from each subject's first level analysis to the second level, where a random effects analysis was performed to produce one-tailed SPM{t} maps. The crossmodal integration conjunction analysis was then performed at the second level using a one-way ANOVA to statistically test for voxels that demonstrated both significant visual-auditory interactions and bimodal elevations (see section 3.1.3-data analysis IV, for details).

3.3 Experiment 1 results: visual-auditory object recognition

The primary purpose of this experiment was to identify where in the brain visual and auditory sensory perturbations arising from the same environmental events are crossmodally integrated to produce unified bimodal perception. Unimodal contrasts were also performed to establish whether significant voxels from the crossmodal integration conjunction were located within, or beyond, unimodal sensory cortex.

3.3.1 Experiment 1 unimodal subtraction analyses

Subjects were exposed to both unimodal visual and unimodal auditory components of the same video clips, providing the opportunity to isolate brain areas involved in processing dynamic sensory events. To control for low-level sensory processing in the crossmodal integration conjunction, pure visual white noise was presented during unimodal auditory stimulus epochs and pure auditory white noise was presented during unimodal visual stimulus epochs. This means that the unimodal contrasts should reveal only those intermediate to high-level brain areas involved in detecting, extrapolating and recognising stimulus features embedded in the 'noisy' background and not primary sensory areas.

Unimodal contrasts: visual versus auditory (V>A)

The unimodal visual versus auditory contrast (V > A) revealed a large swathe of significant voxels ($p<0.001$, uncorrected) distributed bilaterally on the lateral occipital cortices, spreading posterodorsally into the parietal cortex and anteroventrally along the fusiform gyri (figure 10). Activation of V1 and V2 appeared to be absent, presumably because the pure visual white noise included with all auditory stimuli induced the same level of activity as the noise degraded visual stimuli at these areas - along the calcarine sulci and at the occipital poles. The lateral and ventral occipital activations include the territory of several visual brain areas known to be involved in processing form, movement and objects, i.e. V3A, V5 and the lateral occipital complex (LOC), respectively (Malach et al., 1995; Sunaert et al., 1999; Zeki et al., 2003). The activity at the border between parietal and occipital cortices may reflect processes involved in extracting structure from motion (Paradis et al., 2000), which is the primary source of information available for visual object recognition in this study. Other studies investigating the neural correlates of perceived 2D and 3D structure from motion have

observed the involvement of certain sub-regions of the intraparietal sulcus (IPS) (Vanduffel et al., 2002), which may account for the IPS activity in this experiment. However, it is well established that posterior regions of the parietal cortex are involved in sustained visual attention (Corbetta and Shulman, 2002) and so would be expected to contribute to regional brain activity in the vicinity of the IPS.

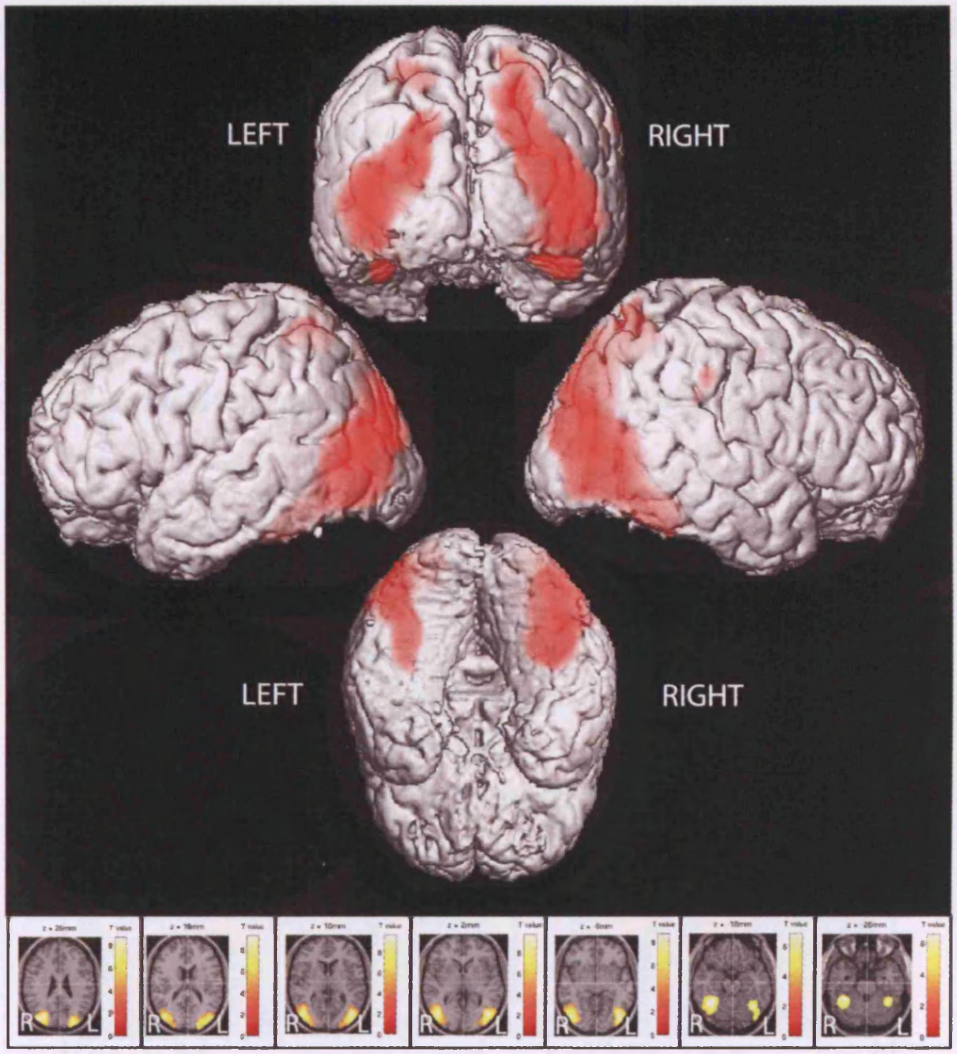


Figure 10: Experiment 1 unimodal visual versus auditory contrast. Significant voxels ($p<0.001$, uncorrected) have been projected onto the surface of a normalised, rendered, structural brain with cerebellum removed (top panel) and in descending transverse sections of the canonical SPM brain (lower panels). Large voxel clusters spread bilaterally throughout ventral and lateral occipital areas, excluding V1 / V2, but extending across the occipito-parietal border into posterior parietal cortex.

Unimodal contrasts: auditory versus visual (A>V)

The unimodal auditory versus visual contrast revealed a large area of significant voxels ($p<0.001$, uncorrected) along the superior temporal gyrus (STG), bilaterally and the left inferior frontal gyrus (IFG) (figure 11). Extensive bilateral activation of Heschl's gyrus (HG) was observed (figure 11Aiv), including primary auditory cortex, which is usually found in the medial two thirds of HG (Rademacher et al., 1993). This contrast also revealed activity in voxels posterolateral to HG in extraprimary auditory regions of the planum temporale (figure 11Aiii,iv), which are known to play an important role in processing spectrotemporal acoustic features (Binder et al., 2000; Giraud et al., 2000; Hall, 2002). Auditory activity rostral and lateral to HG, in the planum polare, superior temporal gyrus and superior temporal sulcus, was particularly extensive in the left hemisphere (figure 11, Aiv,v,vi / B-LEFT). Auditory stimuli also induced extensive activation of the left IFG, which has been identified in previous imaging studies of auditory object recognition (e.g. Maeder et al., 2001).

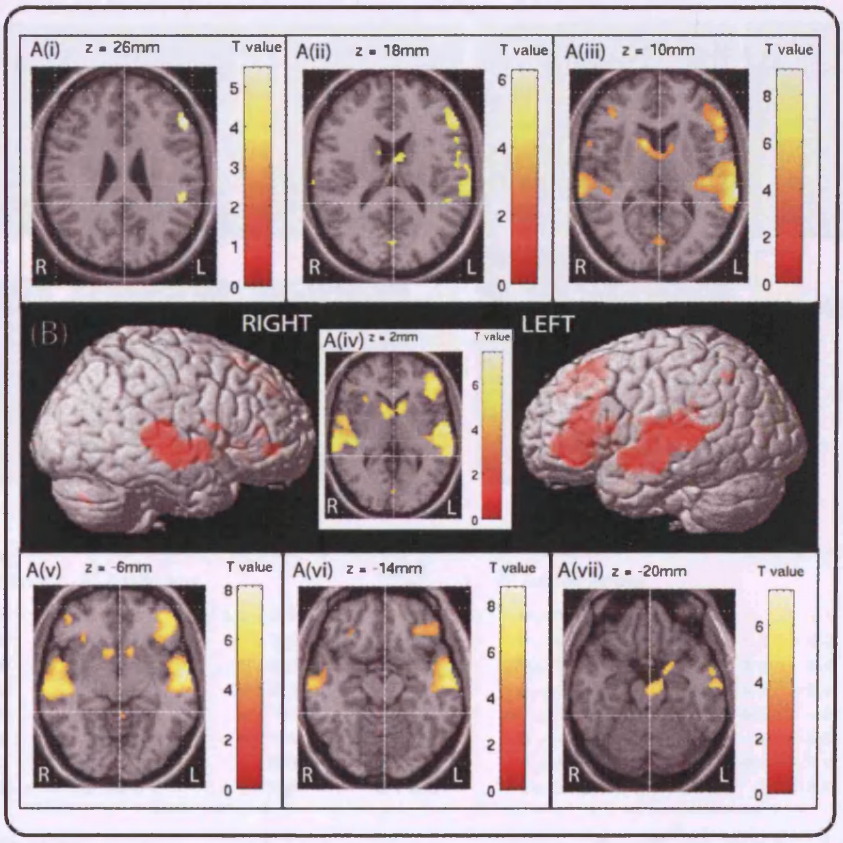
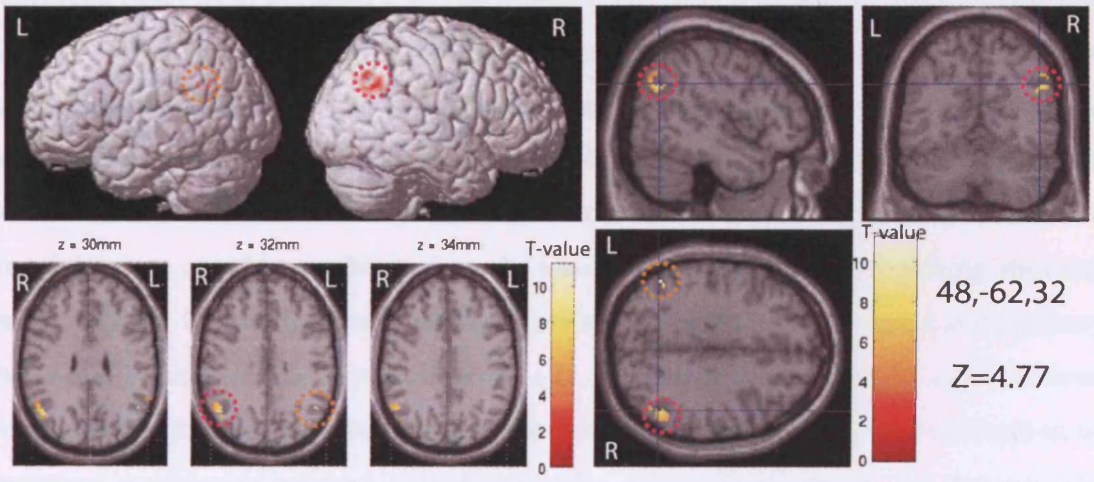


Figure 11: Experiment 1 unimodal auditory versus visual contrast. This revealed significant ($p<0.001$, uncorrected) activations overlaid on descending transverse structural slices (Ai-vii) and surface rendered to the side views of the canonical SPM brain (panel B).

3.3.2 Experiment 1 crossmodal interaction analysis

The visual-auditory interaction and crossmodal elevation contrast for correctly identified stimuli only, comprised the crossmodal integration (CMI) conjunction. This revealed significant ($p<0.001^2$, uncorrected) interactions between concurrent visual and auditory stimulus processing and an elevation in BOLD response magnitude during bimodal as compared to unimodal stimulus conditions, at the temporo-parietal junction (TPJ) bilaterally (figure 12). The TPJ is located at the junction between temporal and parietal lobes, including the posterior-most extension of the superior temporal gyrus and ventral inferior parietal lobe (IPL). This right side activation will be referred to as 'posterior' TPJ, as it lies posterior to the terminal dorsal deflection of the lateral sulcus into the IPL (figure 12, upper panel, side views). The left TPJ activation consisted of fewer suprathreshold voxels, but was found to be statistically more significant (Z=5.18) than its counterpart situated in the right hemisphere (Z=4.77).



voxel-level					x,y,z {mm}		
$p_{FWE-corr}$	$p_{FDR-corr}$	T	$(Z_=)$	$p_{uncorrected}$			
0.000	0.000	11.27	5.18	0.000	-50	-60	32
0.001	0.000	9.11	4.77	0.000	48	-62	32
0.083	0.001	5.66	3.80	0.000	16	66	20
0.271	0.002	4.88	3.49	0.000	-62	-52	26
0.325	0.003	4.77	3.44	0.000	62	-4	-28
0.399	0.003	4.64	3.39	0.000	-6	-52	40
0.571	0.004	4.43	3.29	0.001	-44	-70	38
1.000	0.006	4.09	3.12	0.001	2	-60	34

Figure 12: Experiment 1 crossmodal integration conjunction. Significant voxels of the bilateral temporoparietal junction have been surface rendered on the SPM canonical brain (top left), overlaid on brain sections (top right) and slices (middle left). Coordinates in the results table (bottom) are colour-coded to match the dashed circles outlining the activations to which they refer.

3.4 Experiment 2 methods: visual-auditory synchronicity

3.4.1 Experiment 2 aims

The goal of this experiment was to determine which brain areas are involved in the *regulation* of crossmodal integration, rather than the site of the actual integration itself. In the introduction it was argued, on a purely theoretical basis, that brain areas responsible for determining whether or not a certain pair of visual and auditory stimuli should be integrated in perception, must be critically sensitive to bimodal synchronicity (section 1.2.2). In this study, such areas were functionally distinguished by statistically contrasting regional brain activity during synchronous versus asynchronous, abstract, visual-auditory stimulation.

Previous studies investigating the neural correlates of crossmodal synchrony and asynchrony using functional imaging techniques have invariably elected to use discrete, brief and instantaneously occurring visual and auditory events (Bushara et al., 2001), or transitions (Calvert et al., 2001). In this experiment, fundamentally different visual and auditory stimuli were used that varied gradually between two extremes in a continuous, rather than discrete, manner. This enabled the relative synchrony or asynchrony to evolve over a period of several seconds, rather than occurring suddenly. The motivation for this was that, in previous experiments, synchrony or asynchrony between *discrete*, bimodal events could be established on the basis of matching or non-matching stimulus *onsets / offsets*. In the experimental paradigm devised for this study, visual and auditory onsets and offsets were always concurrent (i.e. in both synchronous and asynchronous conditions), with the synchronicity judgement instead depending upon identification of consistent *covariation* between bimodal stimulus *cycles*. Furthermore, the dynamically varying nature of these stimuli was, in many ways, more faithful to the characteristics of complex bimodal stimuli commonly encountered in normal human environments in that they did not appear and disappear, as in previous imaging studies.

3.4.2 Experiment 2 subjects

Six subjects were recruited, three male and three female, all right-handed and of 21-28 years of age (mean 21.5 years SD ±3.67 years; after one subject was excluded for

sub-standard performance: mean 20.4 years, SD ±2.79 years). These subjects all met with standard selection criteria (section 3.1.2-subject recruitment).

3.4.3 Experiment 2 stimuli

All stimuli in this experiment repeatedly cycled between two extremes in a single stimulus dimension during 12.6s blocks. All stimulus blocks were built from repeats of individual stimulus cycles so single stimulus cycles will be described first (figure 13). This is followed by an explanation of how the temporal regularity of stimulus blocks was manipulated by constructing cycle sequences from same, or different, cycle periods (figure 14). Finally the construction of the bimodal conditions will be explained, which involved pairing like, or unlike, visual and auditory stimulus sequences (figures 15/16).

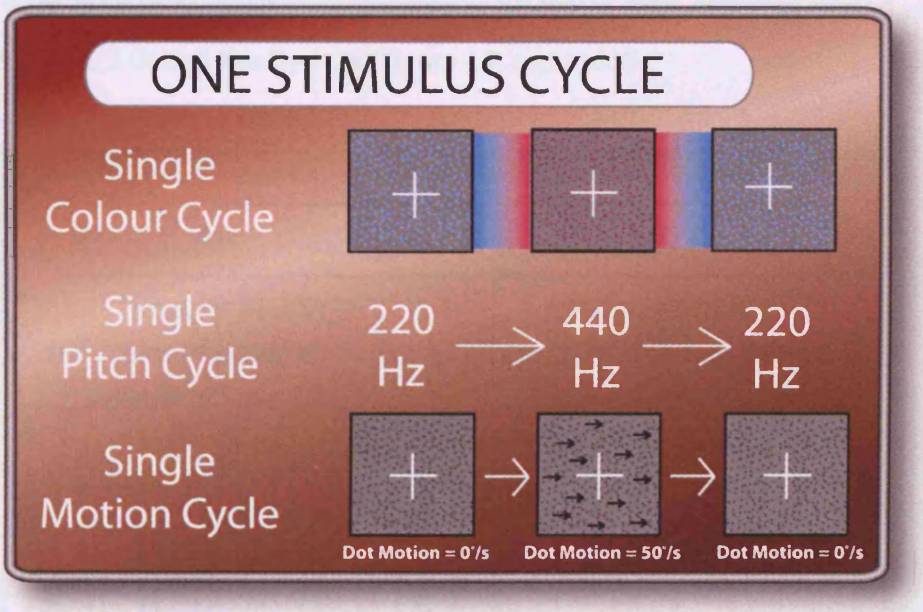


Figure 13: Experiment 2 unimodal stimulus cycles. Visual stimuli consisted either of a stationary, random dot array cycling between cyan and magenta through all intermediate colours (top), or a colourless, moving, random dot array where the unidirectional dot motion cycled between 0°/s and 50°/s through all intermediate velocities (bottom). Acoustic stimuli consisted of a single pure tone cycling gradually between a low pitch of 220Hz and a high pitch of 440Hz (middle).

A single acoustic cycle (figure 13-middle) was a reversing pitch sweep, i.e. the frequency of a single, sinusoidal, pure tone progressed gradually from 220Hz, up to a maximum frequency of 440Hz and back to 220Hz, through all intervening frequencies.

The visual stimulus consisted of a random dot array with 6400 light grey or coloured squares of 0.2° visual angle, on a dark grey background of dimensions 8x8° visual angle. A single visual cycle varied along one of two stimulus dimensions: either the *colour* of static dots cycled continuously from cyan to magenta and back to cyan (figure 13-top), or the unidirectional velocity of colourless (light grey), moving, square, dots steadily accelerated from 0°/s to 50°/s and decelerated back to 0°/s (figure 13-bottom), both progressing through all intervening values.

Manipulating stimulus regularity

For each unimodal stimulus category (i.e. visual-colour, visual-motion or auditory-pitch), subjects were exposed to either regularly, or irregularly, repeating cycle sequences of 12.6s total duration (figure 14). Regular stimuli were created by producing sequences of cycles with a constant period, e.g. all of 2.1s (figure 14, left plot). Irregular stimuli were created by producing sequences of cycles with differing period, e.g. 1.8s, 2.7s, 1.5s, 2.4s, 3.0s and finally 1.2s (figure 14, right plot).

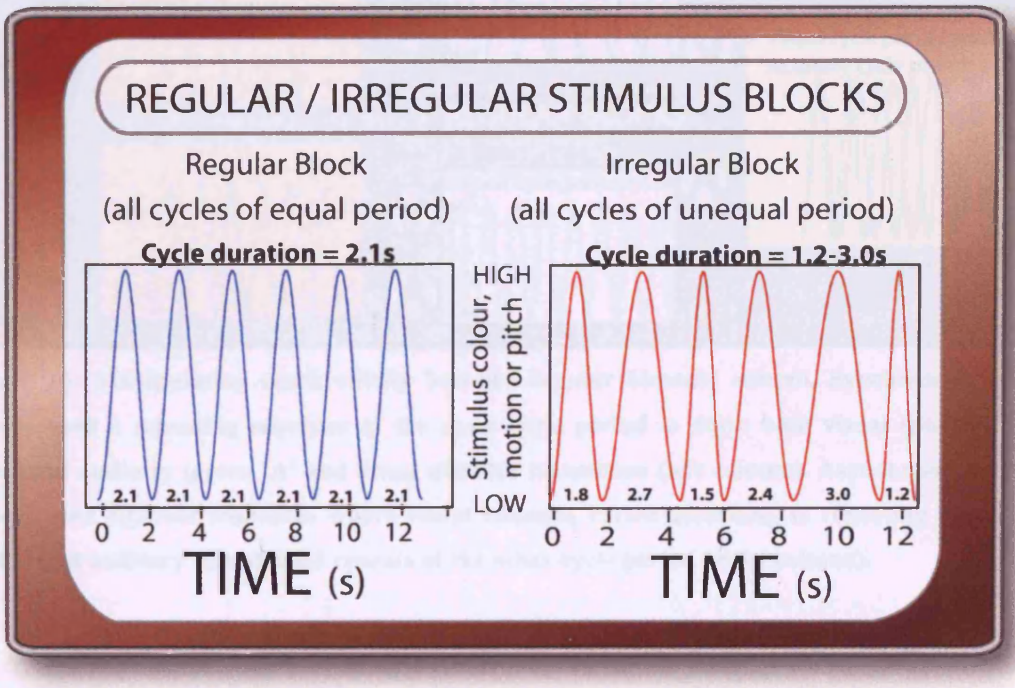


Figure 14: Manipulating stimulus regularity. In regular stimulus blocks the cycle period was kept constant (left graph, values on plot denote period in seconds). During irregular stimulus blocks the cycle period was pseudorandomised through a range from fast cycles of 1.2s duration, to slow cycles of 3.0s (right graph, values on plot denote cycle period in seconds).

Manipulating stimulus synchronicity

The regular unimodal epochs of visual-colour, visual-motion or auditory-pitch transitions either used uniform, repeating cycles either of 1.8s or 2.1s period (figure 15, middle column). To produce *synchronous*, regular, bimodal stimuli, *both* visual and auditory stimuli followed exactly the same cycle sequence, i.e. both sequences were either repeats of 1.8s cycle period or both were repeats of 2.1s cycle period (figure 15, left column). To create asynchronous, regular, bimodal epochs, if the visual sequence consisted of back-to-back repeats following the 2.1s cycle period, then the auditory sequence consisted of consecutive cycles of 1.8s period (figure 15, right column), and *vice versa*. This ensured that there was no consistent crossmodal phasic relationship, despite the fact that both visual and acoustic stimuli start and finish concurrently.

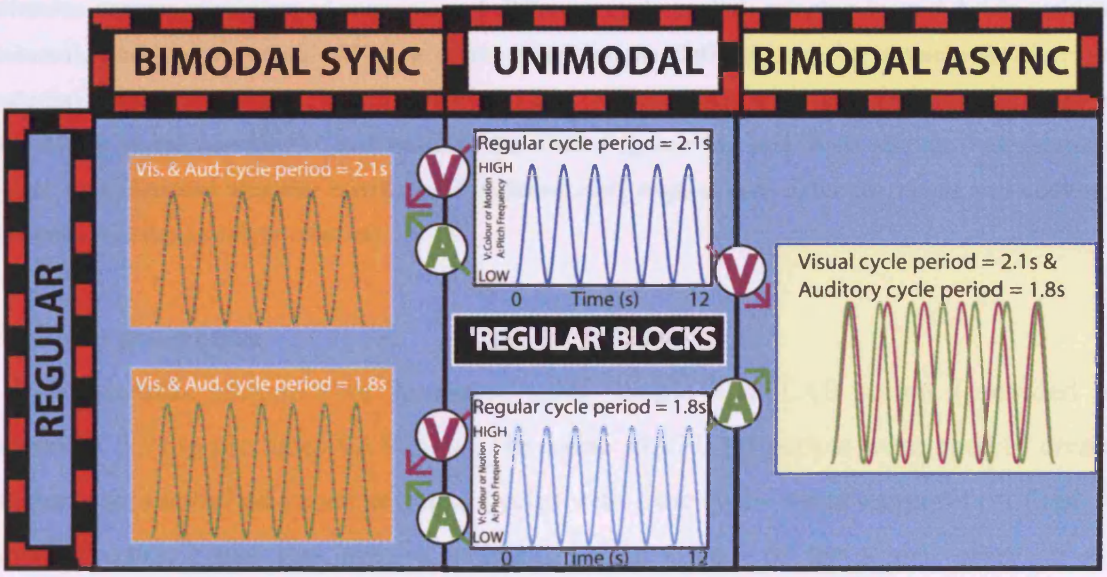


Figure 15: Manipulating synchronicity between regular bimodal stimuli. Synchronous stimulus epochs used a repeating sequence of the *same* cycle period to drive both visual (purple 'V' and lines) and auditory (green 'A' and lines) stimulus transitions (left column). Asynchronous stimulus epochs used *different* sequences where visual stimulus varied according to repeating cycles of one period and auditory stimuli used repeats of the other cycle period (right column).

A similar principle was applied in the bimodal pairing of irregular stimuli, in that both visual and auditory stimuli followed *matching* irregular cycle sequences during *synchronous* epochs (figure 16, left column), whilst *asynchronous* conditions were produced using *non-matching* sequences (figure 16, right column).

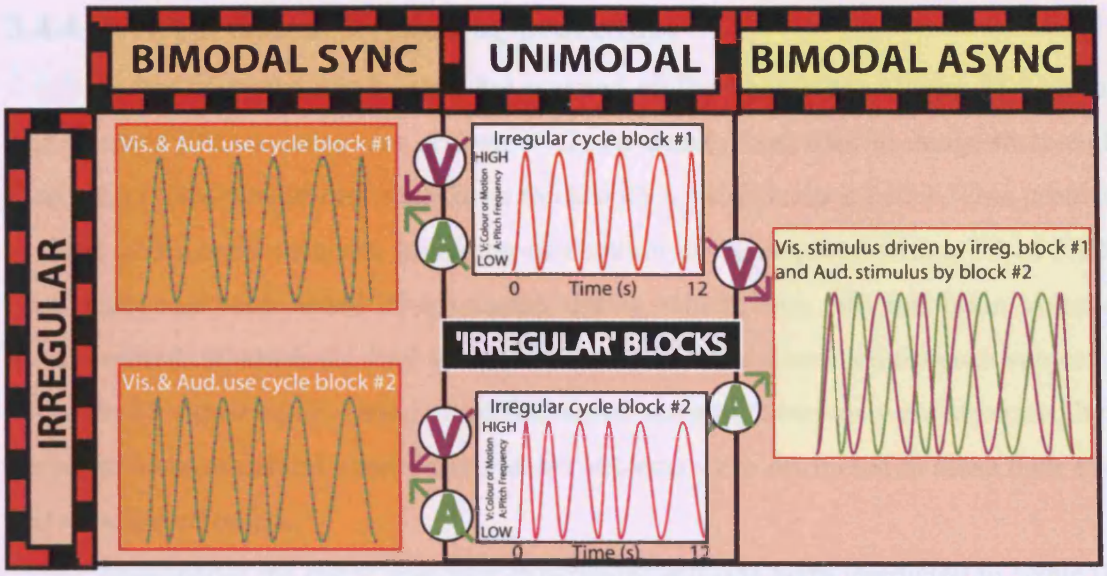


Figure 16: Manipulating synchronicity between irregular bimodal stimuli. Irregular unimodal stimulus vectors consisting of sequences of different cycle periods ranging from 1.2-3.0s (middle column), which were combined to produce synchronous (left column), or asynchronous (right column) bimodal conditions. Synchronous stimulus epochs simply used the *same* cycle sequence to drive both visual (purple 'V' and lines) and auditory (green 'A' and lines) stimuli (left column), whilst asynchronous stimulus epochs used *different* cycle sequences to drive the visual and auditory stimulus transitions (right column).

Stimulus generation

Acoustic stimuli were generated using simple MATLAB scripts (provided in appendix 5.2) to produce WAV files. The same MATLAB scripts were used to create vectors that controlled visual stimulus cycles with exactly the same temporal profiles. A small fixation cross was always overlaid at the centre of the visual array in all conditions.

Furthermore, to ensure that visual and auditory transitions would be properly synchronised when triggered simultaneously, the sample rate used to create the sound files was adjusted according to the refresh rate capabilities of the stimulus PC on which the experiment was to be run. Before each scan session all stimuli were pre-tested to ensure that the program was working properly.

All stimuli were controlled, randomised and coordinated during scanning according to a bespoke Matlab script using COGENT commands. This same script logged subject responses and timing information for each condition, saving a file to disk at the end of each scan.

3.4.4 Experiment 2 scanning procedure

In this experiment subjects were scanned on the Magnetom Allegra 3T scanner. The standard EPI sequence (see section 3.1.2, for details) was used to image forty-eight descending transverse slices, each 2mm thick with a 50% distance factor. This protocol enabled whole brain volumes, including cerebellum, to be acquired with a TR of 3.12s. Each functional scan lasted 21.6 minutes during which time 300 full brain volumes were acquired, of which the first seven dummy scans were discarded for each subject to allow for T1 signal equilibration. Two functional scanning sessions were separated by a twelve minutes structural scan, during which subjects were instructed to close their eyes and remain motionless.

Throughout the entire scanning experiment subjects were instructed to fixate the central cross at all times and to attend to the temporal structure of the stimuli presented. During each scan session subjects were exposed to fifteen different types of stimulus epochs. Six were unimodal stimulus epochs, eight were bimodal epochs and one was a resting baseline condition in which no sound stimulus was played over the headphones and the visual dot array was rendered both colourless and static. Four of the eight bimodal conditions consisted of visual colour cycles paired with auditory pitch cycles and the other four comprised visual motion with auditory pitch transitions. These two bimodal stimulus groups were further subdivided into the regular bimodal or irregular bimodal epochs where *both* the visual and auditory cycle sequences were either regular or irregular. Finally, these four categories of bimodal stimuli could either be rendered synchronous or asynchronous by pairing like or unlike visual cycle sequences with auditory cycle sequences as described above (section 3.4.3).

Subjects were presented with twelve repeats of eight different types of bimodal conditions (2x2x2 factors: synchrony / asynchrony; regularity / irregularity; colour / motion visual submodality), six unimodal conditions (2x3: regular / irregular; colour / motion / pitch) and rest conditions, across two functional scans. Blocks of these fifteen different conditions were presented in pseudo-randomised order, ensuring that the same condition was never presented twice in succession. After exposure to each stimulus the subject was required to indicate, by means of a two alternative forced choice (2-AFC) button press, whether bimodal stimulus transitions were synchronous or asynchronous and, as there can be no synchrony during unimodal epochs, whether unimodal stimulus transitions were regular or irregular.

3.4.5 Experiment 2 data analysis

In order to qualify for inclusion in the analysis, subjects were required to demonstrate a high level of response accuracy - thresholded at a hit rate of at least 90%. All subjects performed in the range of 93 – 100% (mean 96.3% S.D. ± 4.2), with the exception of one subject who achieved a hit rate of 78% (Figure 17). Whilst the other subjects reported that they found the task easy, subject CR stated that they were just not able to do it, so on the basis of substandard performance (>2 S.D. below mean performance) this subject was excluded from the analysis.

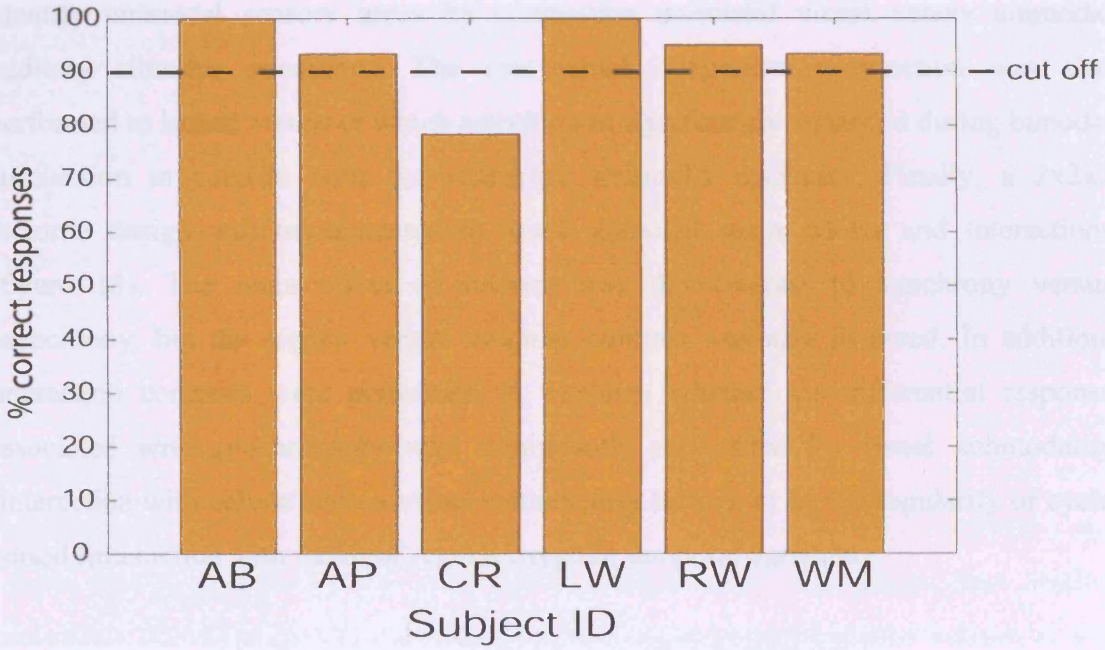


Figure 17: Mean subject performance. Performance was averaged across two sessions and whilst most subjects (n=5/6) found the task very easy, with performance accuracy in excess of the 90% cut-off, one subject was very poor, with almost 25% wrong answers. For this reason subject CR was excluded from the SPM analysis.

Pre-processing for this study included spatial realignment, followed by slice-timing and normalisation to the standard EPI template. These data were then spatially smoothed with a Gaussian kernel of 12mm Full Width Half Maximum and then the pre-processed data were modelled with a boxcar function convolved with a set of basis functions and entered into a multiple linear regression. In order to achieve increased flexibility with regard to the HRF onset time and duration and to better model the variance in the BOLD signal, the boxcar function for each of the fifteen stimulus

conditions was modelled with a regressor not only for the canonical HRF, but also its time derivative and dispersion derivative. For this small group of subjects, analyses were carried out at the group fixed effects level using F-contrasts (Friston et al., 1999). As F-tests are two-tailed, in order to establish which of the contrasted conditions induced the greater activation, it was necessary to plot the parameter estimates.

In this experiment all results presented have been statistically thresholded at $p<0.05$, corrected for multiple comparisons, unless otherwise stated. The known functionally specialised visual areas selectively responsive to unimodal colour and motion stimulation, namely areas V4 and V5 respectively, were first isolated to verify that the data had been suitably modelled. Simple subtraction contrasts were also used to identify unimodal sensory areas by contrasting unimodal visual versus unimodal auditory stimulus conditions. The crossmodal integration conjunction was also performed to isolate voxels at which activity was significantly enhanced during bimodal stimulation in contrast with the respective unimodal responses. Finally, a 2x2x2 factorial design was implemented to assess bimodal main effects and interactions (figure 18). The main effect of interest was the contrast of synchrony versus asynchrony, but the regular versus irregular contrast was also assessed. In addition, interaction contrasts were performed to establish whether the differential response associated with synchronicity was significantly modulated by visual submodality (interaction with colour/motion visual submodality factor), or by the regularity of cycle period (interaction with factor of regular/irregular temporal variation).

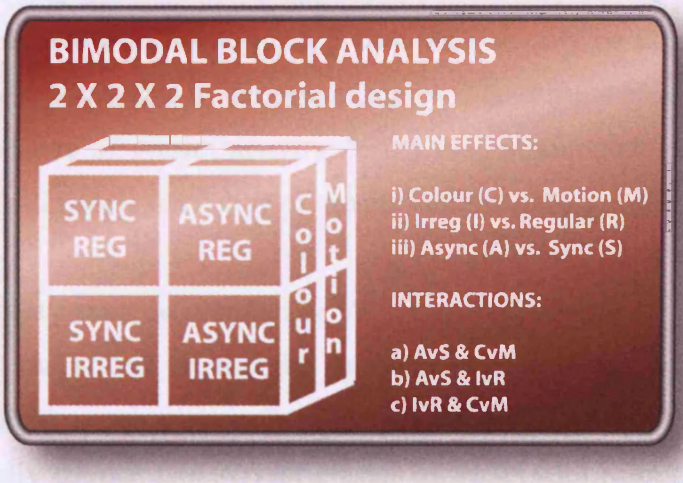


Figure 18: Experiment 2 factorial design. Diagram depicting the 2x2x2 factorial design used to interrogate the data with regard to bimodal main effects of interest.

3.5 Experiment 2 results: visual-auditory synchronicity

3.5.1 Experiment 2 unimodal subtraction analyses

In order to ensure that the data had been properly modelled by the chosen analysis unimodal visual motion cycles were contrasted with unimodal visual colour cycles. If motion-related activity could be detected in the human visual motion area (V5) and colour-related activity in the human visual colour area (V4), these contrasts would confirm that the analysis was set up properly.

As each condition was modelled by three, separate, orthogonal regressors (the haemodynamic response function (HRF), time derivative and dispersion derivative), F-contrasts were used, as this approach enables a linear sum of all three parameter estimates to explain the data rather than individual parameter estimates. Of these regressors, the parameter estimate for the HRF is the most important, as it reflects response magnitude, whilst the time and dispersion derivatives merely allow the canonical HRF additional flexibility in terms of its onset time and overall duration. It is important to bear in mind that F-contrasts are two-tailed and so yield voxels at which the activation profile is significantly *different* under one condition versus another. In other words, taking the contrast of unimodal visual motion (M) versus unimodal visual colour (C) as an example (MvC), the F-contrast will reveal both voxels that were significantly more active during motion stimulation than colour stimulation (M>C) *and* voxels that were significantly more active during colour stimulation than motion stimulation (C>M). Which of the stimulus conditions induced the greater activity at any of the voxels rendered significant by an F-contrast, can only be ascertained by plotting the HRF parameter estimates. In the case of the M-C contrast, a positive HRF parameter estimate is indicative of greater activity during motion stimulation, whilst a negative HRF parameter estimate is indicative of greater activity during colour stimulation.

Visual submodality contrasts: isolating V5

SPM{F} maps were created for the contrast of unimodal visual motion versus unimodal visual colour stimulation (M>C), enabling the human motion area (V5) to be isolated bilaterally (figure 19). A vast cluster of 5132 significant ($p<0.05$, corrected) voxels was centred around the medial occipital pole (maximally significant voxel: [-2,-88,-2, MNI]) including the foveal confluence and spreading laterally to include the right

V5 activation (local maxima: [46,-72,4, MNI]) residing in the lower few transverse slices (figure 19Ai: z=2 to z=6). Left V5 was clearly delineated with a distinct cluster of 183 significant voxels [-46,-74,10, MNI], centred in the upper few transverse slices (figure 19Ai: z=8 to z=12). The positive HRF parameter estimates, for both right (figure 19Aii) and left (figure 19Aiii) V5, confirmed that activity was significantly greater during unimodal motion than colour conditions. Had the F-contrast been performed in the reverse direction (i.e. C-M), the HRF parameter estimate would have been of exactly the same magnitude, but a negative rather than positive deflection. This would have led to precisely the same inference, as both reflect lesser activity during colour than motion conditions, i.e. the greater activity during exposure to moving than coloured stimuli.

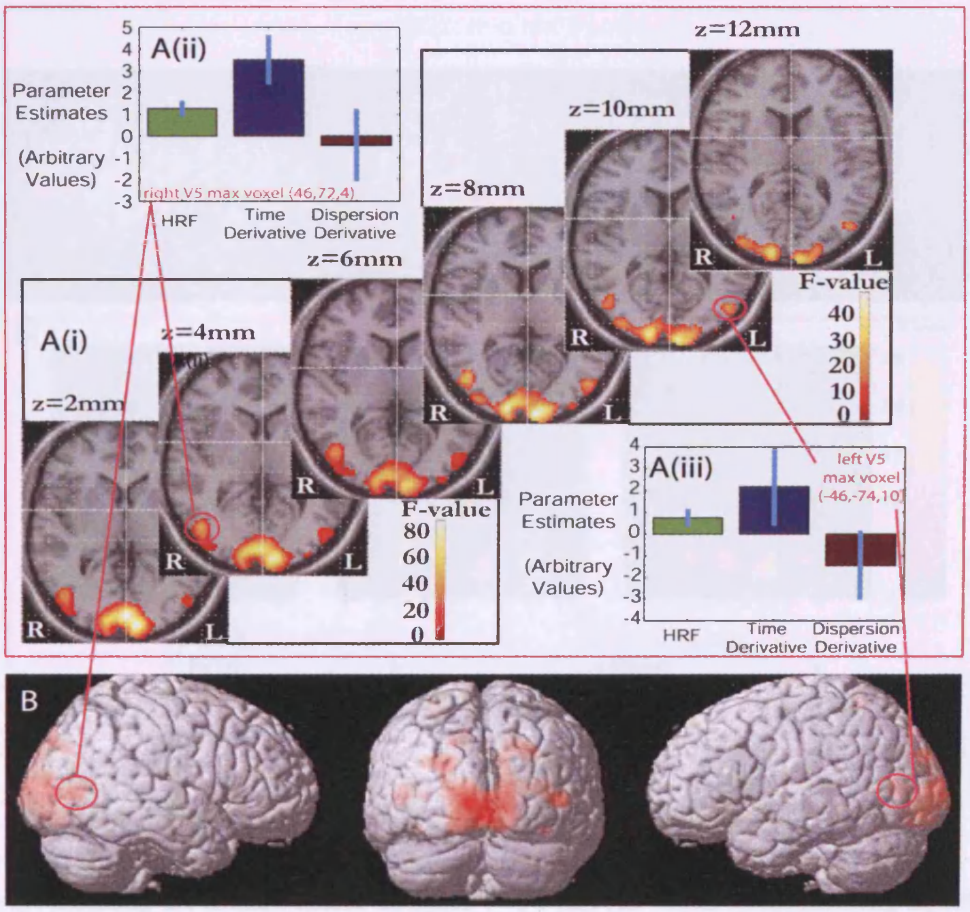


Figure 19: Isolating the human motion area (V5). Contrasting visual unimodal moving versus coloured stimulation (i.e. M>C) reveals bilateral V5 activations overlaid on ascending transverse slices of the canonical SPM brain (Ai). The HRF parameter estimates demonstrate greater activity during motion than colour stimulation at both right V5 (Aii) and left V5 (Aiii) voxels. All significant ($p<0.05$, corrected) voxels were also surface rendered on the canonical SPM brain (B).

Visual submodality contrasts: isolating V4

Figure 20A once again illustrates all significant voxels from the MvC contrast surface rendered on the canonical SPM2 brain, but here the position of the V4 activations have been highlighted. Bilateral activation of the V4 complex is also illustrated with voxels overlaid on transverse slices through a representative subject's (subject WM) normalised structural image (figure 20Bi). The negative HRF parameter estimates for both the maximally significant voxel in right V4α [30,-50,-20, MNI] (figure 20Bii) and in left V4 [-28,-66,-18, MNI] (figure 20Biii), indicates significantly lesser activity during unimodal motion than unimodal colour stimulus conditions; i.e. these voxels were more active during colour stimulation. These coordinates correspond to those previously reported for V4 and V4α (Bartels and Zeki, 2000).

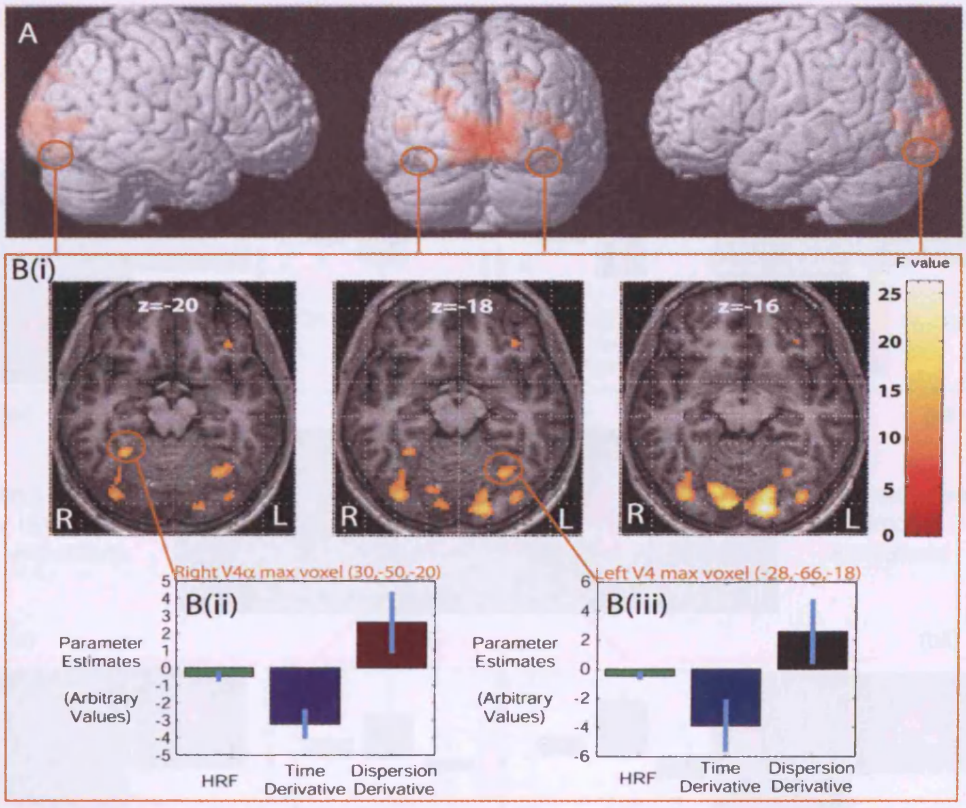


Figure 20: Isolating the human colour complex (V4/V4α). All voxels demonstrating significantly ($p<0.05$, corrected) different activation profiles during visual unimodal motion versus colour stimulation (i.e. M>C) were surface-rendered on the canonical SPM2 brain (A). Voxels corresponding to V4 and V4α have been overlaid on ascending transverse slices of subject WM's normalised structural brain (Bi). Negative HRF parameter estimates for right V4α (Bii) and left V4 (Biii) voxels indicate greater activity during visual unimodal colour than motion stimulation.

Unimodal contrasts: visual versus auditory (V > A)

The visual versus auditory contrast revealed significantly different responses throughout visual and auditory sensory cortices. These included voxel clusters distributed bilaterally along the superior temporal gyri, both ventral and medial regions of occipital cortex and at the caudal occipito-parietal border (figure 21, central panel). The maximally significant voxel clusters were located in the superior temporal plane (figure 21a). The negative deflection of the HRF parameter estimates indicates a larger BOLD response during auditory than visual stimulation (figure 21ai & aii). The bilateral activations of the superior temporal plane include parts of the planum temporale on the left and parts of the planum polare on the right (figure 21a, transverse sections). Visual activations were revealed at medial, ventral and superior occipital cortex (figure 21, central panel) with the most significant voxels in the caudoventral occipital cortex bilaterally. Positive HRF parameter estimates at these loci confirm that they were more active during unimodal visual than auditory stimulation (figure 21bi & bii).

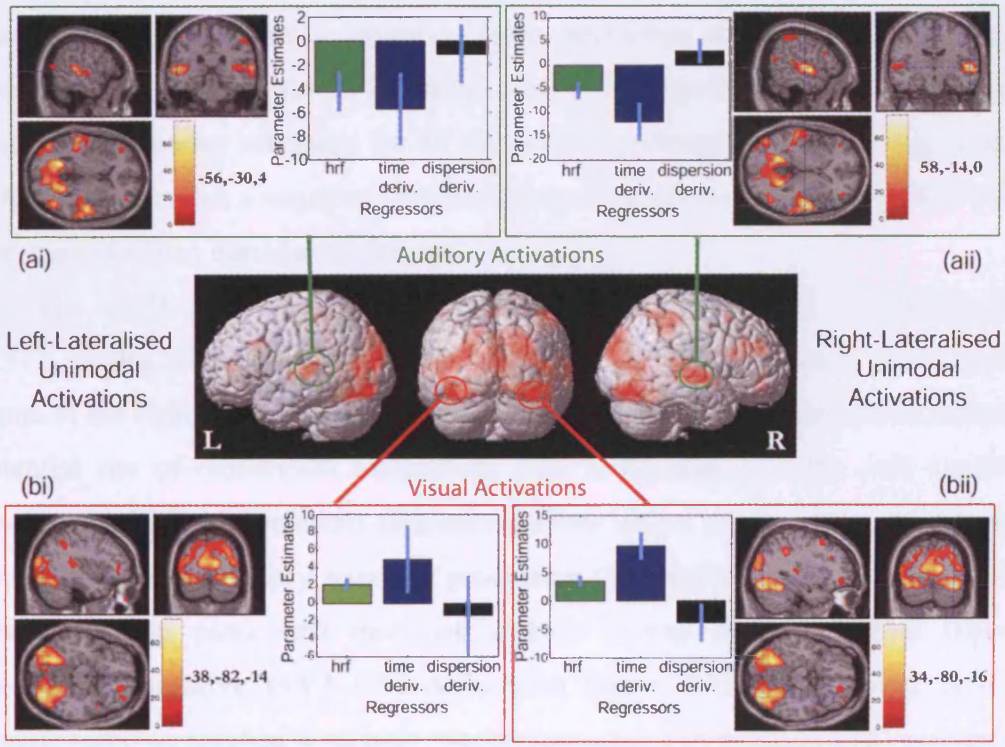
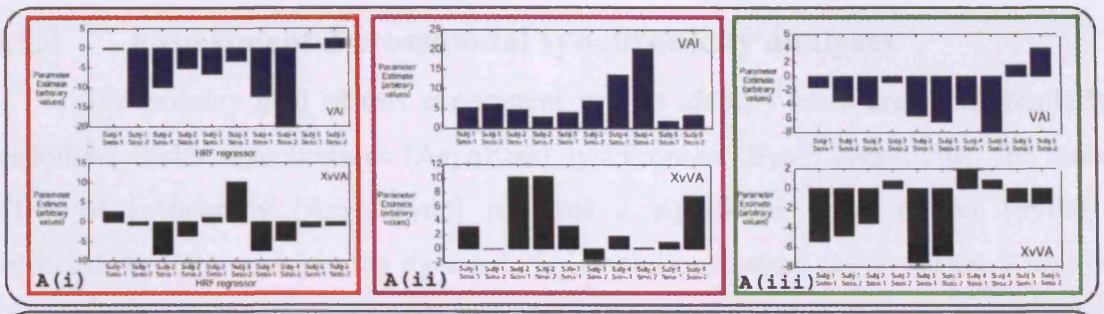


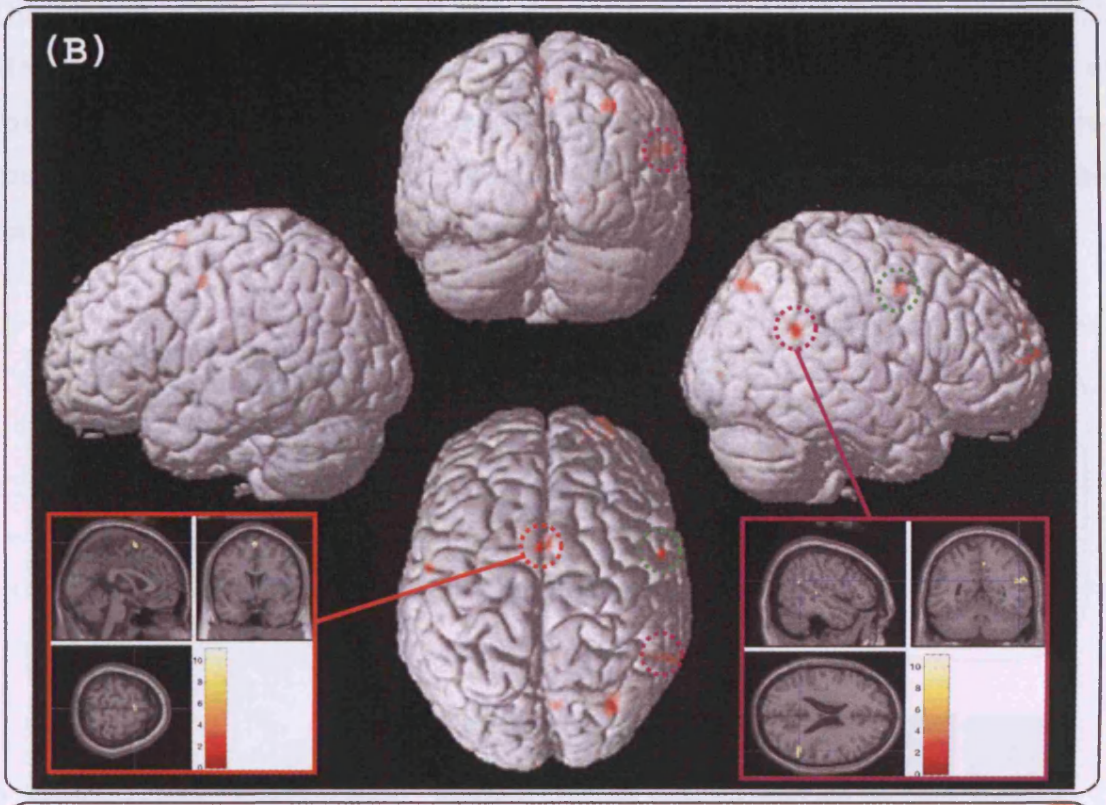
Figure 21: Experiment 2 unimodal contrasts. Contrasting unimodal visual versus auditory responses revealed significantly different responses ($p<0.05$, corrected) throughout visual and auditory sensory areas. Negative deflections in the HRF parameter estimates (green bar) at voxels in bilateral auditory cortex (ai-left & aii-right) indicate greater responses to auditory than visual stimulation. Greater activity during visual than auditory stimulation is reflected in the positive deflections of the HRF parameter estimates at voxels in bilateral visual cortex (bi-left & bii-right).

3.5.2 Experiment 2 crossmodal integration analysis

The crossmodal integration (CMI) conjunction was performed for the asynchronous and synchronous bimodal conditions separately, but resulted in essentially the same distribution of significant ($p<0.001^2$, uncorrected) clusters. When the CMI conjunction was performed on all stimuli together, voxels at several brain loci were found to be statistically significant ($P<0.05^2$, corrected) (figure 22). The parameter estimates for the three most significant activations were plotted (figure 22A). (NB in figure 22A only the HRF parameter estimates have been plotted for each subject individual scanning sessions (i.e. 5x2 = 10) demonstrating the session-to-session consistency of these effects). These activations have been surface rendered on the canonical SPM brain (figure 22B) and tabulated (figure 22C) using a system of colour-coding to identify correspondence between parameter estimates, location of the activation on the surface of the rendered structural image and row of the results table. The HRF parameter estimates demonstrate that, of the statistically significant activations, only in the right anterior temporo-parietal junction (aTPJ) does the BOLD response reflect both *positive* visual-auditory interaction effects and is significantly *increased* during bimodal versus unimodal conditions (figure 22Aii). The negative load on the HRF parameter estimates for all the other significant activations (e.g. figure Ai and Aiii) indicates that a negative visual-auditory interaction and greater BOLD activity under unimodal than bimodal conditions.

The aTPJ activation at the maximally significant voxel [52,-50,22, MNI; Z=7.31], resides within a cluster of 59 suprathreshold voxels located at the posterior extreme of the right superior temporal gyrus. The aTPJ is the only activation considered a potential site of crossmodal integration. This is because it is the only significant activation at which the parameter estimates indicate both a positive factorial interaction between visual and auditory stimulus processing (i.e. positive VAI HRF deflections, figure 22Aii top plot), and increased activity during bimodal versus unimodal processing (i.e. positive XvVA HRF deflections, figure 22Aii, bottom plot). However, this same response resulted from both synchronous and asynchronous CMI conjunctions (data not shown). As the right aTPJ activations appeared to demonstrate no preference for synchronous bimodal stimuli over asynchronous bimodal stimuli it seems unlikely that it is a true site of crossmodal integration for bimodal perception (a more likely explanation is discussed in section 4.3).





(C)

cluster-level	voxel-level						
k_E	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	F	$(Z)_{\equiv}$	$p_{uncorrected}$	x,y,z {mm}	
50	0.000	0.000	11.00	7.32	0.000	0	4 68
59	0.000	0.000	11.00	7.31	0.000	52	-50 22
	0.000	0.000	9.27	6.61	0.000	62	-52 24
17	0.000	0.000	10.28	7.03	0.000	56	0 44
46	0.000	0.000	10.12	6.97	0.000	34	-76 46
22	0.000	0.000	9.15	6.57	0.000	30	60 8
6	0.000	0.000	9.12	6.55	0.000	52	-28 4
18	0.000	0.000	8.79	6.41	0.000	-10	-72 28
12	0.000	0.000	8.55	6.31	0.000	4	-52 48
11	0.000	0.000	8.48	6.27	0.000	6	-74 52
12	0.000	0.000	8.44	6.26	0.000	-56	-6 46

Figure 22: Experiment 2 crossmodal integration conjunction. Subject and session-specific HRF parameter estimates have been plotted for the three most significant activations from the crossmodal integration conjunction (A). All significant ($p<0.05^2$, corrected) voxel clusters are surface rendered on the canonical SPM brain (B) and tabulated (C). Like activations in A, B and C are colour-coded accordingly. Of the three most significant activations, only the right anterior temporo-parietal junction demonstrated positive HRF deflections in both contrasts (Aii) indicating both positive interaction between visual and auditory processing and bimodal response elevations.

3.5.3 Experiment 2 crossmodal synchronicity analyses

The primary goal of this experiment was to identify brain areas differentially responsive under asynchronous (Asyn) and synchronous (Sync) conditions. The main effect of asynchrony [Asyn>Sync] revealed a significant voxel cluster ($p<0.05$, corrected; k=100 voxels) in the right anterior frontal operculum (aFO) (figure 23ai-iii). This activation comprised a small sub-cluster deep in the Sylvian fissure [36,18,2, MNI], straddling both opercular and anterior insula territories. A continuous body of voxels, spreading along the inward-facing anterior frontal opercular tissue in a ventro-lateral trajectory, connected the deep sub-cluster to a larger superficial cluster containing the maximally significant voxel [48,22,-8, MNI] (figure 23aiii).

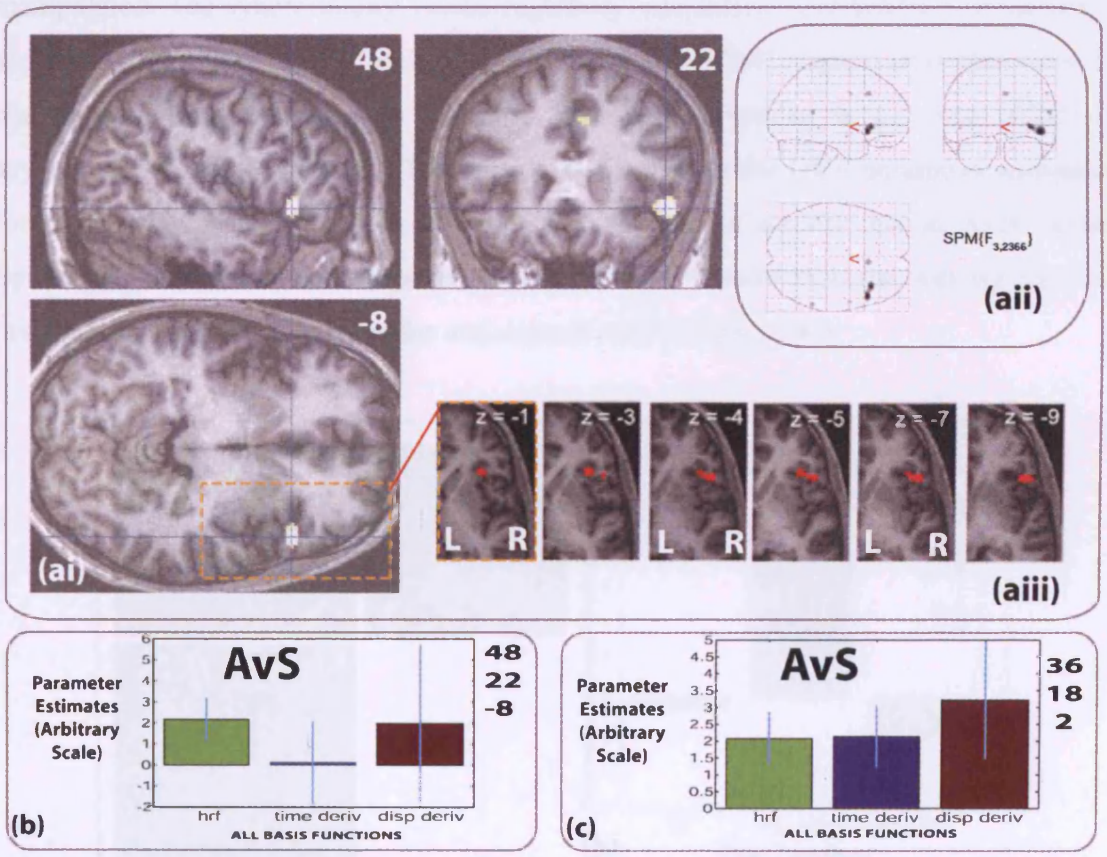


Figure 23: Contrasting asynchrony versus synchrony. This contrast reveals a cluster of 100 significant voxels ($p<0.05$, corrected) at the right aFO overlaid on sections of a representative subject's normalised structural image (ai), projected into a glass brain (aii) and onto descending transverse slices (aiii). Parameter estimates for outer (b) and inner (c) local maximally significant voxels are shown in the bottom panels. The positive deflection of the HRF regressor parameter estimates indicates greater activity occurred under asynchronous than synchronous conditions.

The positive deflection of the HRF parameter estimates plotted for both the superficial (figures 23b) and deep (figure 23c) sub-clusters is indicative of greater activity under Asyn than Sync conditions. Contrary to expectations, no voxels were identified that were more active during synchronous versus asynchronous conditions. The other main effect of interest was the regularity contrast (IvR), which yielded no significant voxels at corrected significance at all.

A secondary aim of this analysis was to determine whether significant interactions might be identified between the factor of primary interest - crossmodal synchronicity - and the other factors: namely, colour versus motion defined visual cycles and regular versus irregular cycles. No significant results were obtained in the synchronicity x visual submodality interaction, suggesting that the neural correlates of asynchrony detection were not modulated by the specific visual submodality that was manipulated. The synchronicity versus regularity interaction, [Asyn-Sync]-[Irreg-Reg], yielded a significant cluster of 33 voxels [36,26,-12, MNI] (figure 24a) that were in rostroventral juxtaposition with the deep subcluster identified in the main effect of asynchrony versus synchrony. The negative deflection of the HRF parameter estimates for the interaction contrast indicates a greater increase in activity during Asyn versus Sync stimulation when the temporal structure of the bimodal stimulus was regular and predictable, as opposed to irregular and unpredictable (figure 24b).

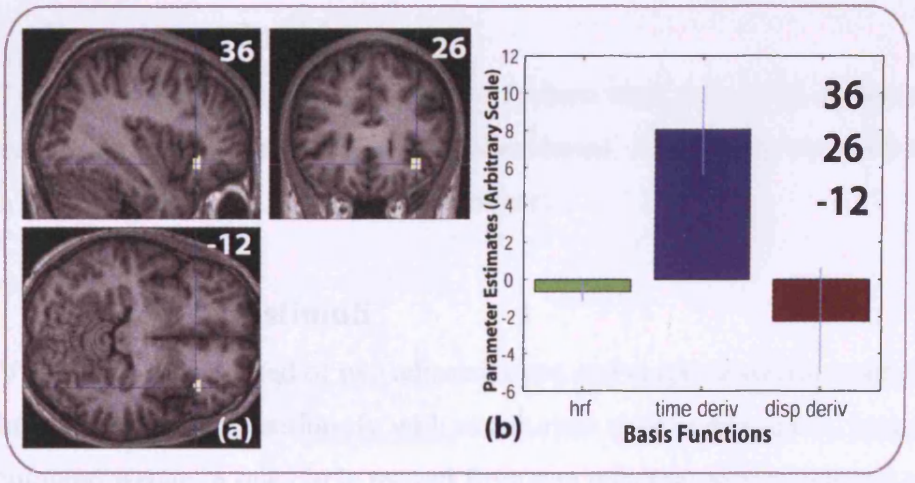


Figure 24: Factorial interaction between synchronicity and regularity. A significant ($p<0.05$, corrected) functional interaction between synchronicity and regularity was identified in the deep right anterior frontal operculum overlaid on structural sections of a representative subject's normalised brain (a). Negative deflection of HRF parameter estimate for this contrast indicates significantly less asynchrony related activity for irregular versus regular bimodal stimuli (b).

3.6 Experiment 3 methods: cross- and intra-modal synchronicity

3.6.1 Experiment 3 aims

Having identified that the right anterior frontal operculum was more active during crossmodally asynchronous than synchronous conditions, responses to intramodal asynchrony (i.e. asynchrony between two visual stimuli and asynchrony between two auditory stimuli) were also investigated in this study. The aim was to establish whether heightened BOLD activity at this locus was induced only by crossmodal asynchrony or also by intramodal asynchrony. This necessitated a stimulus paradigm where two visual and two auditory stimuli were presented simultaneously in all epochs. The timing of discrete, cyclic transitions within each of these four 'channels' was pseudo-randomised according to the Poisson distribution, to produce total temporal independence and therefore asynchrony between the four channels. Synchrony was introduced between different channels to selectively manipulate crossmodal and intra-modal temporal covariation as required. The neural correlates of crossmodal synchronicity could thus be distinguished from those of intra-modal synchronicity by interrogating the data according to a simple 2 x 2 factorial design.

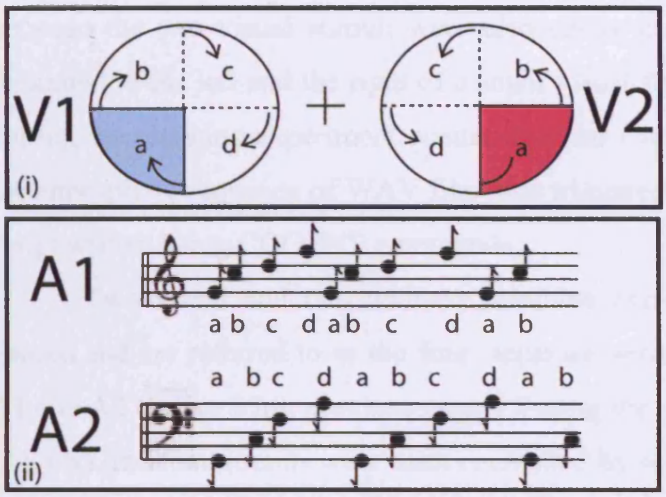
3.6.2 Experiment 3 subjects

Twelve subjects were recruited, seven of whom were male, with an age-range of 19-27 years (mean=24.3 years ± 2.3 standard deviation). All subjects met with standard selection criteria (section 3.1.3-subject recruitment).

3.6.3 Experiment 3 stimuli

Visual stimuli consisted of two adjacent cyan and magenta circles, each with one of the four quadrants filled uniformly with either cyan or magenta pixels, respectively. As the coloured wedge in one circle moved from one quadrant to the adjacent quadrant in a clockwise direction, the other would move from position to position in the same four-phase cycle, but in an anticlockwise direction (figure 25i). Auditory stimuli consisted of two sets of four, single, musical instrument notes, played sequentially in order of ascending pitch. When the highest pitch note was reached the sequence started

again, thus progressing cyclically through the four auditory phases in a manner analogous to the four visual phases (figure 25ii).



(iii) Inter-Stimulus Intervals (s)

PHASE	V1	V2	A1	A2
a	0	0	0	0
b	1.12	0.32	0.64	0.80
c	0.72	1.36	1.12	0.56
d	1.04	0.64	1.04	0.96
a	0.40	0.48	0.32	0.96
b	0.88	0.48	0.64	0.88
c	0.72	0.72	0.88	1.36
d	1.20	1.04	1.04	1.12
a	0.96	0.56	0.96	0.88
b	0.80	0.96	0.72	1.28

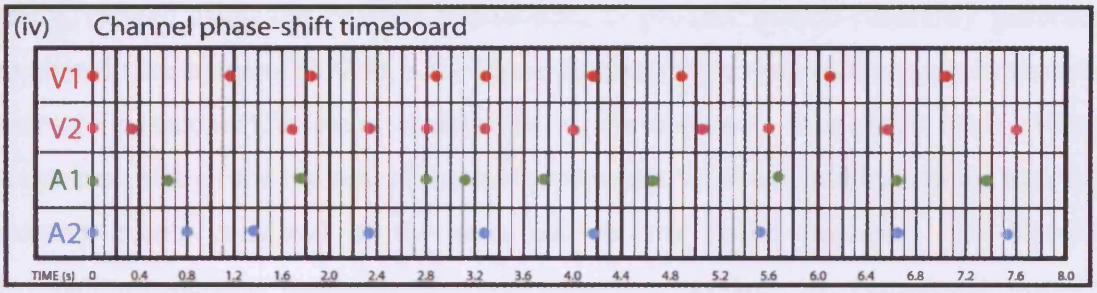


Figure 25: Experiment 3 stimuli. The four phases (a→b→c→d) of the two visual (i) and two auditory (ii) stimulus cycles are represented in the top left panels. Each stimulus channel (i.e. V1, V2, A1 and A2) was ascribed a single column vector from the matrix of pseudorandomly generated inter-stimulus intervals, where successive rows describe the interval between each successive stimulus phase for each of the four independent channels (iii). This particular example resulted in a combined stimulus array in which all four stimulus channels progressed asynchronously with respect to one another, which is best visualised in the 'channel phase-shift timeboard' (iv). In order to introduce synchrony between stimuli the selected channels were simply assigned the same, rather than independent, inter-stimulus interval vectors.

The two sets of sounds were created using a synthesiser / keyboard (ROLAND Phantom XA, UK) to produce instrumental notes of vastly differing sound qualities. One set was high-pitched, with rapid rise (attack) and fall times and the characteristic timbre of a harpsichord-like instrument. The other set mimicked a deep-pitched, slow rise and fall times and distinctive timbre of a woodwind instrument. Each individual note was saved as a WAV file of 100ms duration. The impetus for choosing auditory stimuli of these descriptions stemmed from the requirement that the two notes should be

as perceptually distinct from each other as possible. By using instrument sounds with entirely different acoustic features, e.g. timbre, register, attack etc, the two sets of notes could be qualitatively distinguished when played simultaneously. Simultaneous changes between the two visual stimuli were also easily distinguished as they were spatially separated to the left and the right of a small visual fixation cross in all stimulus epochs. During the scanning experiment, visual stimulus cycles were entirely controlled by and the appropriate sequence of WAV files was triggered according to a bespoke MATLAB script written using COGENT commands.

Two visual and two auditory stimulus cycles were presented in all stimulus epochs and are referred to as the four, separate, sensory information channels: V1, V2, A1 and A2 (figure 25iii, previous page). During the completely asynchronous condition (ASYN), the four stimuli were each controlled by a unique inter-stimulus interval (ISI) vector, created using the Poisson distribution, to produce pseudo-randomly generated numbers in the range of 0.32 to 1.36. These numbers were then used to govern the time interval (in seconds) between each event in the channel's four-phase cycle. Visual events consisted of the rotation of the coloured wedge from one quadrant to the next and auditory events consisted of the next note in the cyclic sequence. The Poisson distribution-generated randomisation was found to produce the occasional stimulus epochs in which a pair of stimuli seemed subjectively to be synchronous, despite the fact that they were in fact mathematically asynchronous. Constraints were therefore added to the procedure for generating the asynchronous ISI matrix to ensure that events in any one channel were always offset with respect to the other channels by a minimum of 200ms. Hence onset times are described as 'pseudo'-randomised. Therefore, when all four channels were assigned different pseudorandomised ISI vectors, the resulting stimulus array was *entirely asynchronous*, in that events in one channel always occurred at a different time to those of the other channels (figure 25iv, previous page).

Synchrony was introduced between specific channels by assigning them the same ISI vector (e.g. one column from the example ISI matrix in figure 25iii, previous page), whilst the other two channels were assigned a different ISI vector (e.g. a different column from the ISI matrix). This resulted in events that precisely and consistently overlapped in time throughout the stimulus epoch, between the desired channel pairs only. The other channels were temporally independent with respect to the covarying channels. Intramodal synchrony was produced by introducing synchrony between the two visual stimulus cycles and between the two acoustic stimulus cycles (VVAA).

During these conditions there was no crossmodal synchrony (i.e. they were crossmodally asynchronous) as each intramodal pair followed a different ISI vector. In other epochs synchrony was introduced between the two sets of crossmodal pairs (VAVA). During such epochs there was no intramodal synchrony (i.e. stimuli were intramodally asynchronous) as each crossmodal pair was assigned a different ISI vector. The completely synchronous condition (SYNC), in which all stimuli were crossmodally AND intramodally synchronous, was simply produced by having all visual and auditory events controlled by a single ISI vector. This resulted in four different stimulus conditions (figure 26).

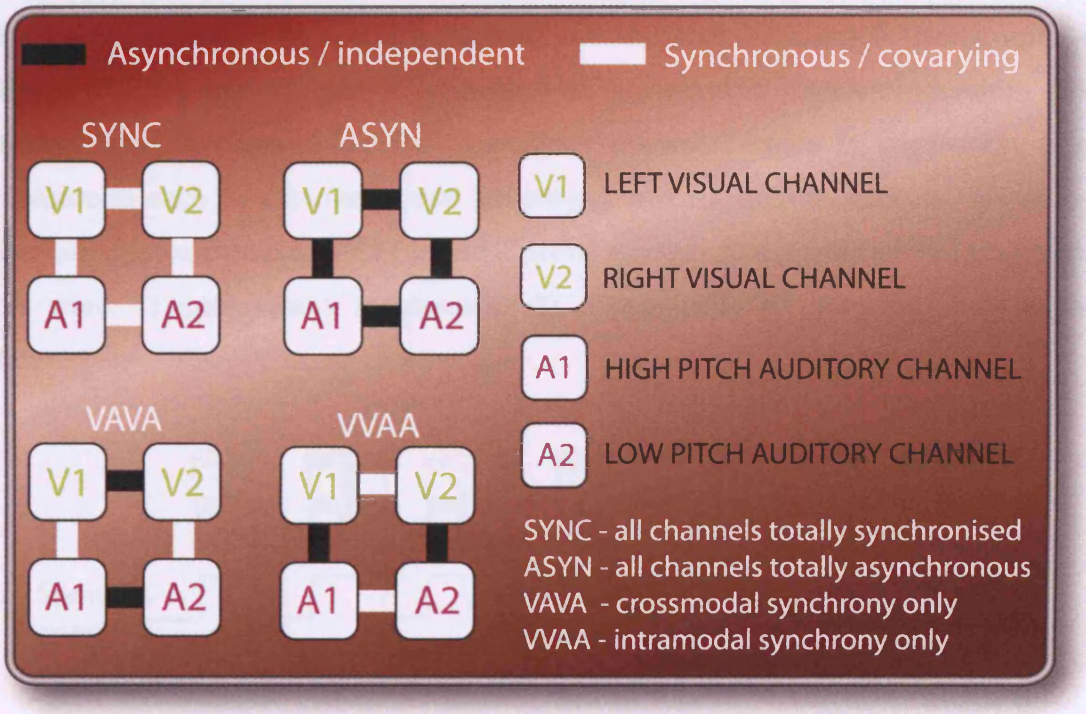


Figure 26: Introducing synchrony between stimuli in experiment 3. Four different types of stimulus epoch were generated, by assigning the same or different ISI vectors to each channel. In the SYNC condition all four channels (V1, V2, A1 and A2) cycled synchronously. During ASYN epochs all four channels cycled asynchronously. For crossmodal synchrony (VAVA), transitions between one crossmodal pair (V1-A1) were driven by one ISI vector, whilst those for the other crossmodal pair (V2-A2) were driven by a different ISI vector. In the final condition (VVAA) visual channels were synchronised to one ISI vector (V1-V2) and auditory channels to another (A1-A2).

3.6.4 Experiment 3 scanning procedure

Subjects were scanned on the Magnetom Allegra 3T scanner (Siemens, Erlangen, Germany) using the standard T2*-weighted, gradient echo-planar imaging (EPI) sequence. Whole brain volumes, including cerebellum, of 32 x 3mm thick, transverse, slices were acquired every 2.08s (TR). All other MRI scanning parameters have been described previously (section 3.1.2-MR scanning information).

Subjects were instructed to fixate the central cross at all times, attend to the two visual and the two auditory stimuli and evaluate the synchronicity of stimulus transitions. Across the two functional scanning sessions subjects were presented with each of the four different epoch categories twenty-four times. These two functional scans were separated by a structural scan and all three each lasted twelve minutes. Blocks of these four different conditions were presented in pseudo-randomised order ensuring that the same condition was never presented twice in succession. After exposure to each 5s stimulus epoch the subject was required to indicate, via a three alternative forced choice (3-AFC) button press, whether it consisted of stimulus cycles that were: 1) intramodally synchronous 2) crossmodally synchronous or 3) neither (figure 27).

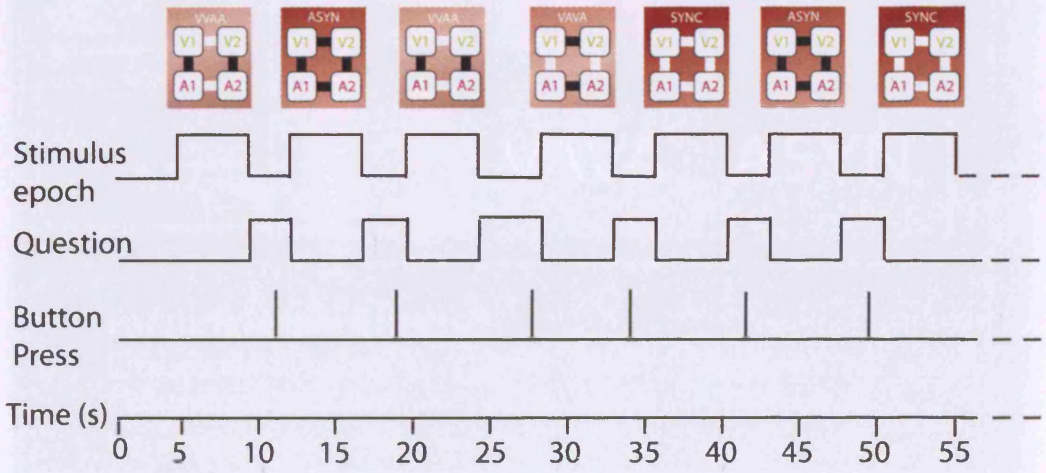


Figure 27: Experiment 3 stimulus presentation paradigm. Onsets and offsets of the pseudo-randomised 5s 'stimulus epoch' alternated with the fixed 2s 'question' epoch, during which subjects were required to decide whether the stimuli were intramodally or crossmodally synchronous or neither, and register their selection with an appropriate button press.

3.6.5 Experiment 3 data analysis

Pre-processing consisted of spatial realignment, normalisation to standard stereotactic coordinates and smoothing (see section 3.1.3-data analysis I).

Each stimulus condition was modelled by a single boxcar function convolved with the canonical haemodynamic response function and entered into a multiple linear regression. Statistical contrasts were set up to investigate the main effects of the 2x2 factorial design (see figure 28), which was implemented in order to investigate differences in regional brain activity during *crossmodal* asynchrony versus synchrony, and *intramodal* asynchrony versus synchrony. Random effects analyses were carried out by performing separate one-sample t-tests at the second level on the CON images that resulted from the subject-specific subtraction analyses at the first level.

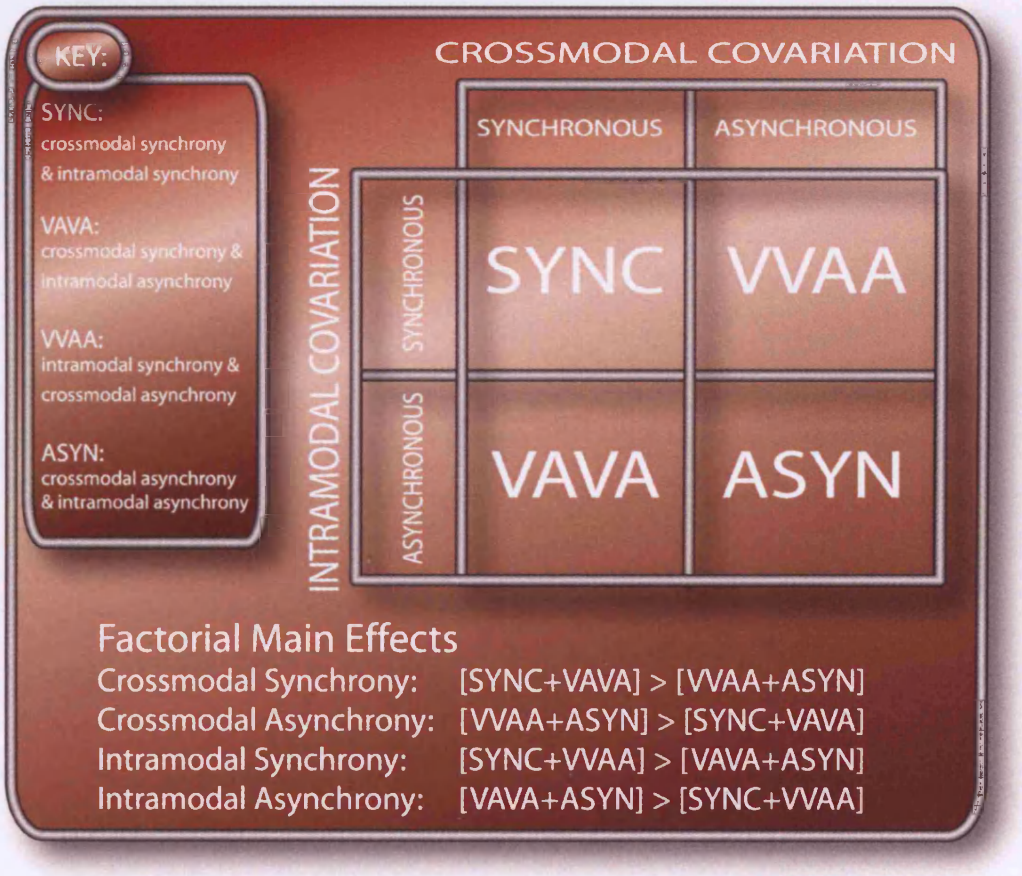


Figure 28: Factorial design for experiment 3. Crossmodal and intramodal synchronicity could be manipulated independently enabling a factorial analysis. Each cell of the factorial design diagram contains a code indicating the pairs of channels that were synchronised (i.e. driven by the same ISI vector). The other channels were asynchronous with respect to the synchronised channels (see key for full explanation of individual conditions). The main effects of interest were contrasted as described below the 2x2 factorial grid to which it relates.

Unlike the two-tailed SPM{F) maps used in the previous experiment, the SPM{t} maps used in this experiment are one-tailed. This means that in a contrast between two conditions (A vs B), rather than revealing voxels demonstrating significantly different activity irrespective of which condition caused the greater activity (i.e. A <> B), the SPM{t} maps are directional (A > B), revealing only voxels which demonstrate greater activity during condition A than condition B. This eliminates the need for plotting parameter estimates to investigate which condition induced the greater activity at significant voxels.

3.7 Experiment 3 results: cross- versus intra-modal synchronicity

Having established that the right aFO was more active during asynchronous than synchronous bimodal conditions in Experiment 2, this experiment was devised to investigate whether it was specifically activated by crossmodal asynchrony, or if it would also be activated by asynchrony between stimuli in the same sensory modality (i.e. intramodal asynchrony). This experiment also provided the opportunity to revisit the neural correlates of crossmodal synchrony. Since every condition in this experiment involved bimodal stimulation and subjects were never exposed to unimodal stimuli, it was not possible to perform the CMI conjunction, nor the unimodal contrast analyses.

3.7.1 Main effect of crossmodal asynchrony

The crossmodal asynchrony contrast revealed two large clusters of right-lateralised activity comprising 507 significant (uncorrected, $p<0.001$) voxels at the anterior frontal operculum (aFO) (figure 29A&C) and a separate cluster of 220 voxels located in the dorsal inferior frontal gyrus (IFG) (figure 29B). Activity in the right aFO can be observed in descending transverse slices at z-coordinates 0, -4 and –8mm and includes voxels located within the caudomedially adjacent anterior short insular gyrus (figure 29D), which was also observed in the analogous contrast of experiment 2. Having repeated the observation of increased BOLD activity in the right aFO during crossmodal asynchrony, the findings from experiment 2 have been confirmed. This provides further evidence to support the inference that this region is involved in detecting crossmodal covariation violations.

The right dorsal IFG activation has been reported in previous neuroimaging studies (Strange et al., 2000), where it is referred to as the 'oddball area'. The activity at this site was found to be elevated during deviation from sensory expectations, whether these expectations were perceptual, semantic or emotional in nature. In the 'oddball' experiment all stimuli were visually perceived words, suggesting that this area is not involved in crossmodal asynchrony detection. Instead it appears to respond to any deviation from expectation, irrespective of the modality/modalities through which it was perceived and irrespective of whether these were deviations from low-level sensory or high-level cognitive expectations.

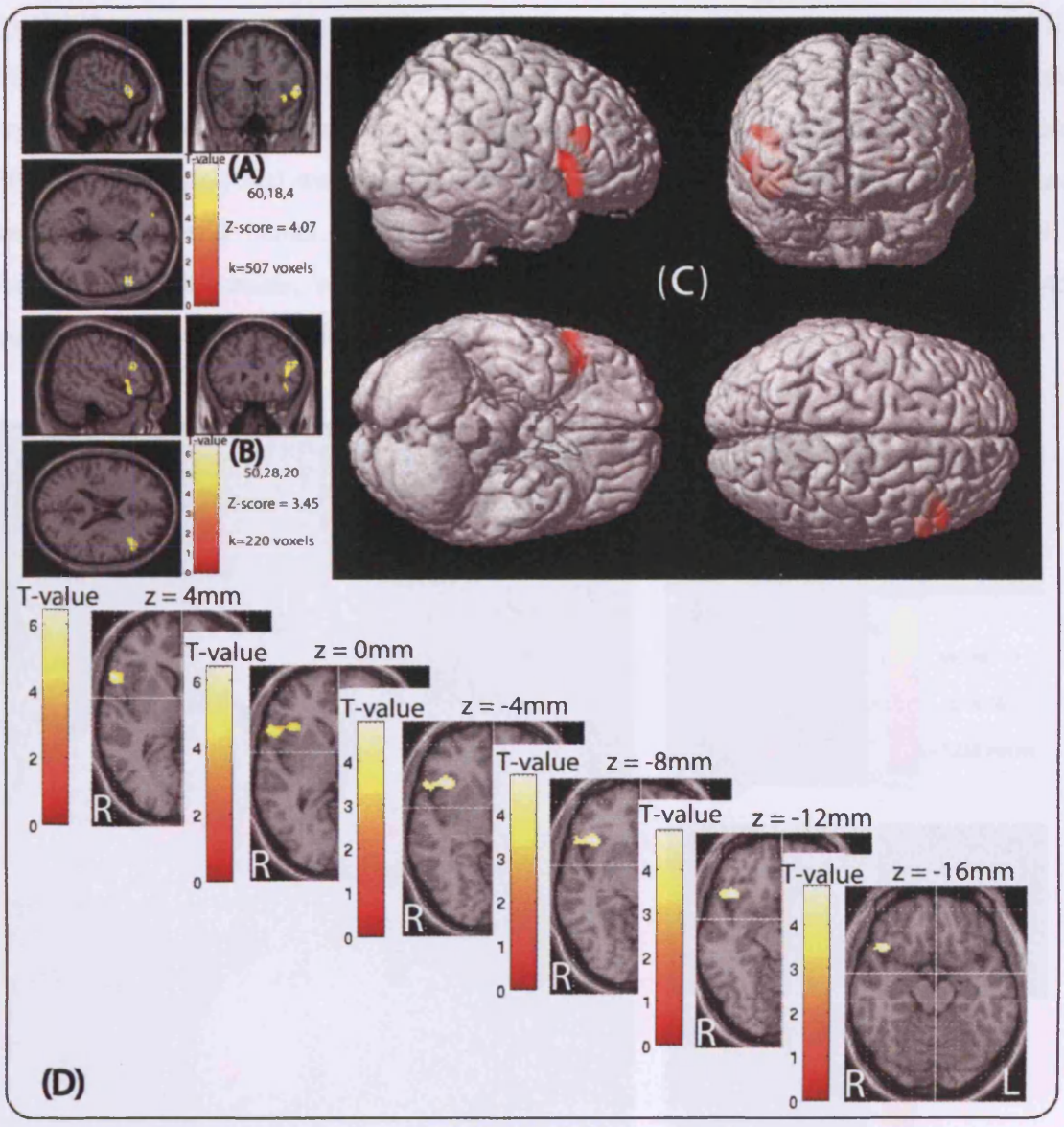


Figure 29: Experiment 3 main effect of crossmodal asynchrony. Significant ($p<0.001$, uncorrected) voxels from the crossmodal asynchrony versus synchrony contrast are overlaid on sections (A and B), surface-rendered (C) and overlaid on descending transverse slices (D) of the canonical SPM structural brain. These included a cluster of 507 contiguous voxels with maximally significant voxel in the right anterior frontal operculum (A) and extending into the adjacent anterior insula (D).

3.7.2 Main effect of intramodal asynchrony

The main aim of this experiment was to establish whether the increases in activation at the right aFO resulted specifically from crossmodal asynchrony, or whether intramodal asynchrony might also induce activations at this site. The results of the intramodal asynchrony versus synchrony contrast (figure 30, left panel) included no significant voxels in the vicinity of the right aFO, instead revealing significantly

($p<0.001$, uncorrected) increased activity along the superior temporal sulcus (STS) bilaterally (figure 30, top right panel) and area V5, bilaterally (figure 30, bottom right panel). The maximally significant voxel was in the right middle temporal gyrus (figure 30, top right panel) and was one of 1228 voxels that appeared to be distributed across ventral and dorsal banks of the STS. The right V5 activation [52,-72,10, MNI] incorporated 265 voxels, whilst the left V5 cluster [-46,-76,12, MNI] comprised 145 voxels (figure 30, bottom-right panel).

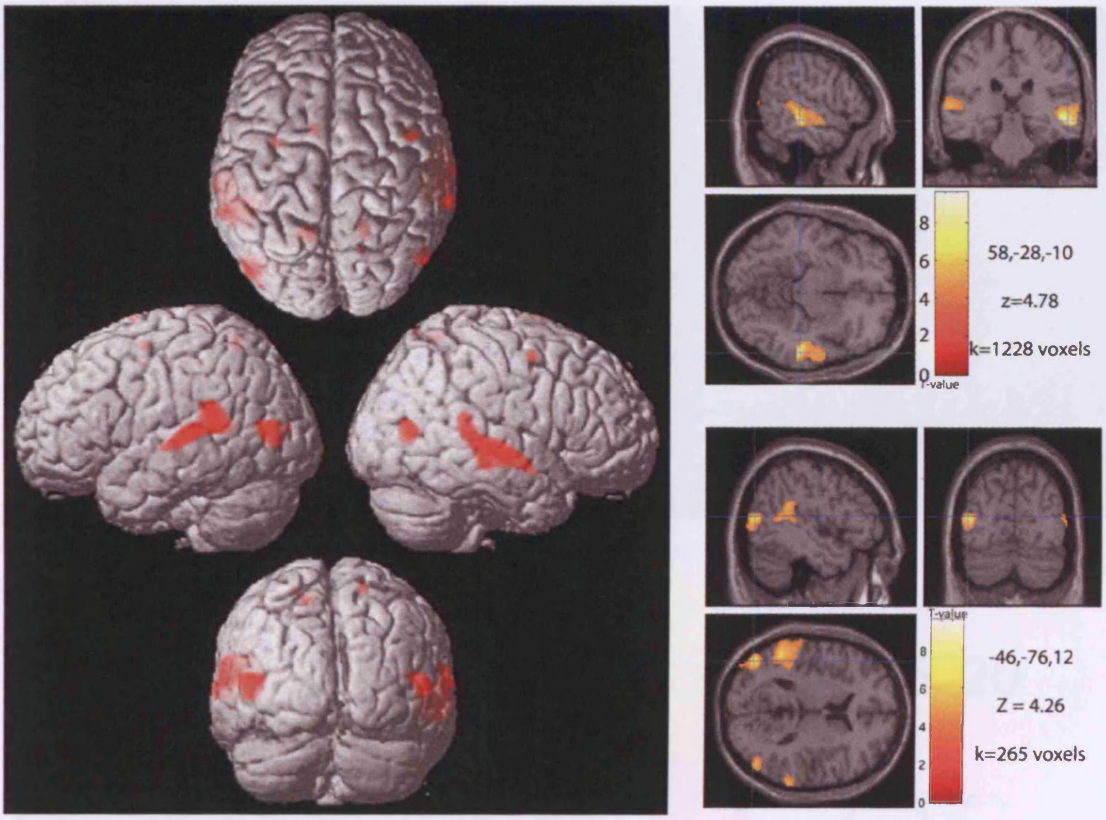


Figure 30: Experiment 3 main effect of intramodal asynchrony. Significant activations in the intramodal asynchrony versus synchrony contrast surface rendered on the canonical brain (left panel). Activations have also been overlaid on structural sections of the canonical SPM brain showing the maximally significant activation in the posterior superior temporal sulcus (top-right panel) and the second most significant voxel within left V5 (bottom-right panel).

3.7.3 Main effect of crossmodal synchrony

The main effect of crossmodal synchrony versus asynchrony revealed significantly ($p<0.001$, uncorrected) increased activity during crossmodally synchronous epochs in a pair of ~200 voxel clusters [±24,-6,-20, MNI] within the amygdala bilaterally (figure 31). No other significant voxels were detected by this

contrast. That these functional activations were located in the amygdala was established on the basis of anatomical criteria (Duvernoy, 1991). Such neuroanatomical references indicate that the amygdala is located ventrally with respect to the globus pallidus, anteromedial to the temporal horns of the lateral ventricles and rostrolateral to the crus cerebri. Furthermore, the coordinates of these bilateral activations closely match those of other recent neuroimaging studies of the amygdala (e.g. [–28,-6,-14, MNI] (Smith et al., 2004) and [–27,-9,-15, MNI] (Gottfried and Dolan, 2004)).

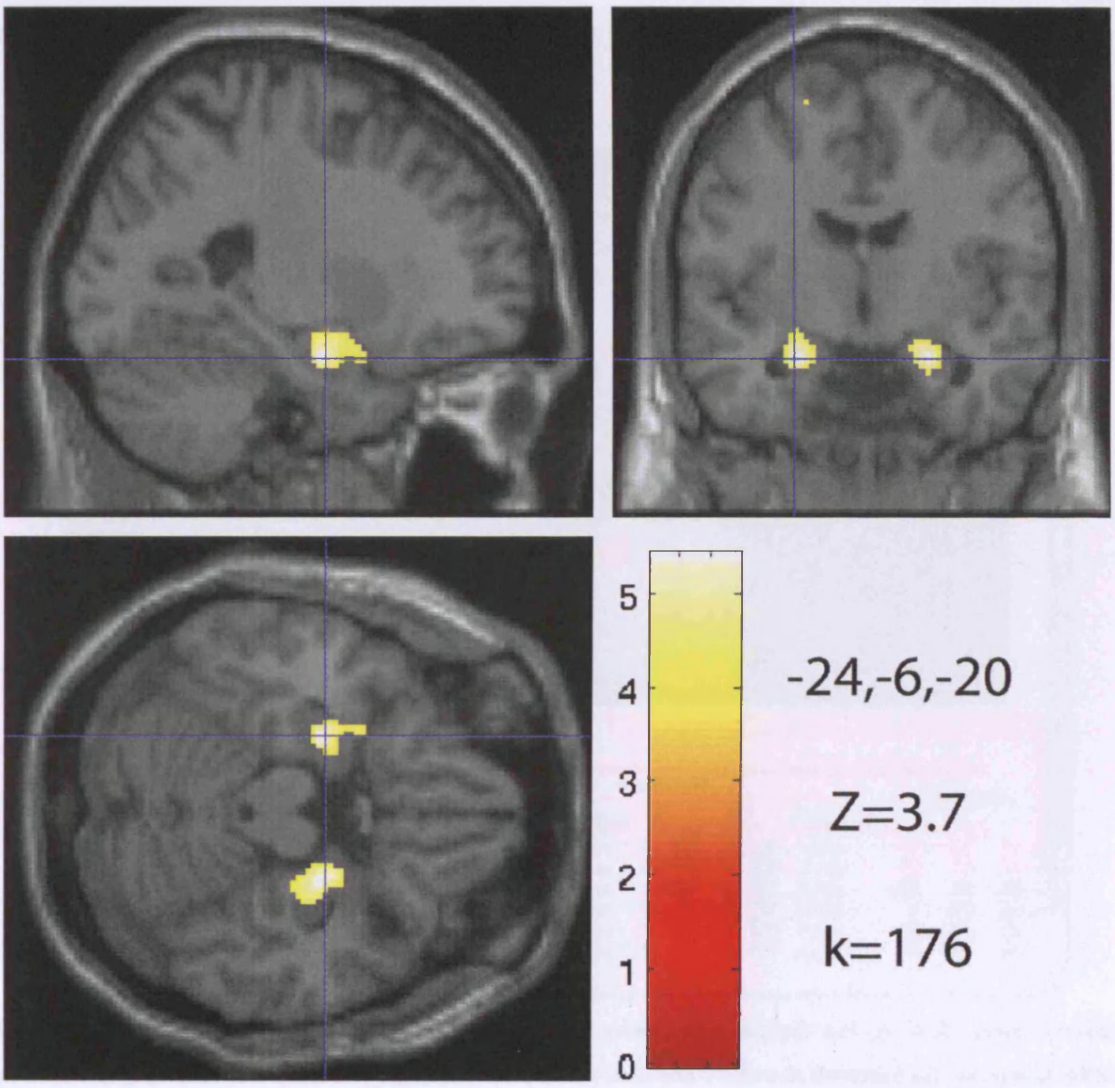


Figure 31: Experiment 3 main effect of crossmodal synchrony. Voxels within the bilateral amygdala overlaid on sections of the canonical SPM brain were significantly ($p<0.001$, uncorrected) more active during crossmodal synchrony than asynchrony.

3.7.4 Main effect of intramodal synchrony

Greater activity during intramodal synchrony versus asynchrony was identified in a pair of large ~900 voxel clusters in the left angular gyrus (figure 32A) and along the parieto-occipital sulcus on the medial surface of the brain (figure 32B). These panels have been colour-coded to match the appropriate colour of coordinates in the results table (figure 32D). The other significant activations have been surface-rendered on the canonical SPM brain (figure 32C) comprising clusters of considerably fewer voxels in the left dorsal inferior frontal gyrus, the posterior superior frontal gyrus, the right angular gyrus and a few small scattered clusters in the superior prefrontal cortex.

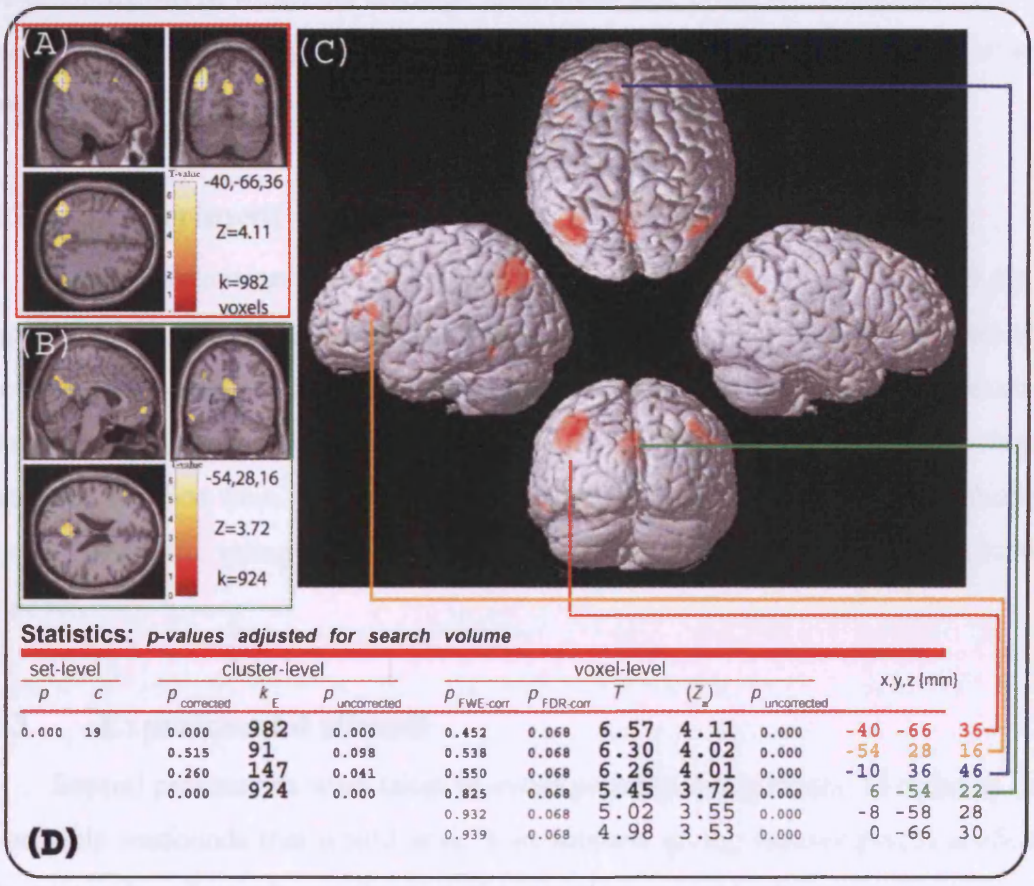


Figure 32: Experiment 3 main effect of intramodal synchrony. Significant (p<0.01, uncorrected) activations from the intramodal synchrony versus asynchrony contrast, overlaid on canonical SPM structural brain sections (A - left parietal; B - occipito-parietal sulcus) and surface rendered (C) on the same brain. The table of results annotates the activations on rendered brain and the voxel coordinates have been colour-coded to match the borders of panels A and B (D).

3.8 Experiment 4 methods: visual-auditory neuroaesthetics

3.8.1 Experiment 4 aims

fMRI was used to scan twelve non-expert subjects who were required to rate a selection of visual art and musical extracts for perceived aesthetic beauty. Each piece was rated individually permitting the identification of the neural correlates of visual and auditory neuroaesthetics. The visual art pieces were also rated during concurrent musical accompaniment and thus differences in regional brain activity under bimodal as opposed to unimodal conditions could be statistically contrasted. Of particular interest were brain regions in which the response magnitude was positively correlated with each subjects' own beauty ratings irrespective of whether ratings pertained to visual artworks or musical extracts, i.e. the neural correlates of supramodal aesthetic appeal.

3.8.2 Experiment 4 subjects

Twelve volunteers (six male, aged 25.1 ± 6.3 S.D.) were recruited as per standard criteria outlined in section 3.1.2 (subject recruitment), with the exception of the following. Individuals with a formal education in art, music or art history were excluded in favour of 'naive' subject populations that were likely to be viewing / hearing the stimuli for the first time. The motivation for this was to minimise the possibility of subjects' aesthetic ratings being influenced by prior knowledge of the artist or composer.

3.8.3 Experiment 4 stimuli

Several precautions were taken to avoid possible rating biases. In order to avoid the possible confounds that would arise from subjects giving famous pieces artificially high ratings based purely on the fact that the piece or the style was familiar, very famous artworks and musical extracts were avoided. Another reason for using relatively obscure pieces was to avoid rating-bias based on subjective nostalgic memory attached to a certain artwork or musical extract.

Forty visual stimuli (listed in appendix 5.3) were digitised on a flatbed scanner (Epson Perfection 1640SU, Seiko Epson Corp, Japan) or selected from online sources (www.abcgallery.com) and saved as tagged image file format (TIFF) images. These

were then manipulated in Photoshop (Adobe, USA) to fill standardised portrait (260 x 300 pixel) or landscape (300 x 260 pixel) dimensions. Any pieces noticeably distorted by this process were excluded from the stimulus pool. A gradual 'zoom' effect was created using Premiere video editing software (Adobe, USA), with the aim of making visual epochs as engaging as the continuously evolving auditory music epochs, which were both of 12s duration. These video clips were rendered as AVI files which is the preferred format for video in Matlab scripts using COGENT commands.

Forty auditory music segments were selected from the classical and jazz music genres (listed in appendix 5.3), digitised from compact disc to hard disc using ACE software (freeware: www.mp3-ripper.com). The Audacity software package (www.sourceforge.net) was used to edit extracts to the required length, normalise the amplification of each segment to avoid clipping and export the final WAV file to be played over the pneumatic headphones during the experiment. The electrostatic headphone system was not available on the 1.5T scanner.

Bimodal stimuli comprised set pairs of visual art and musical extracts rendered as AVI movie files using Premiere. All stimuli were presented and responses collected using COGENT commands implemented in Matlab scripts.

3.8.4 Experiment 4 scanning procedure

This was the only experiment performed using the 1.5 Tesla, Sonata, Magnetic Resonance scanner (Siemens, Erlangen, Germany). Functional scanning consisted of two fifteen minute runs of the experimental stimuli, separated by a structural scan, during which subjects were requested to shut their eyes and relax. Each functional run consisted of 300 full brain volumes, acquired using the standard T2*-weighted, echo planar imaging (EPI) sequence (details in section 3.1.2). In this particular experiment whole brain functional images were acquired, including the entire cerebrum and cerebellum, using 48 transverse slices, with a slice thickness of 2mm, resulting in a TR of 4320ms.

During the first functional scanning session, subjects were exposed to unimodal visual stimuli (V), unimodal auditory (A), bimodal (VA) and rest conditions (R), in pseudorandomised order, ensuring that all subjects were exposed to stimuli in different orders and that no condition was presented twice in succession. During 'V' conditions, subjects viewed one of a pool of twenty visual art pieces, presented as a movie file that

gradually 'zoomed in' to the centre of the image, with no accompanying auditory stimulation. One of a pool of twenty musical extracts was relayed to the subject via the headphones during 'A' conditions, with no accompanying visual stimulation. During 'VA' conditions, subjects were presented concurrently with one of twenty set pairs of visual art and music extracts from a separate stimulus pool. During 'R' conditions subjects were presented with a grey screen and no auditory stimulation apart from the ongoing ambient noise from the MRI scanner. At the end of each 12s stimulus epoch, subjects were invited to indicate their aesthetic beauty rating on a scale of one to eight, via an appropriate button press (figure 33).

In the second functional scan subjects were exposed to all the stimuli previously presented as bimodal pairs under unimodal conditions and all stimuli previously presented under unimodal conditions were presented bimodally. This ensured that all stimuli were rated under both unimodal and bimodal conditions. Whether any particular stimuli were presented separately in the first scan and in combination during the second (and *vice versa)* was counterbalanced across subjects. This ensured that for every subject that rated a particular stimulus pair under unimodal conditions in the first run and under bimodal conditions in the second, another subject was exposed to that same stimulus pair in the opposite order.

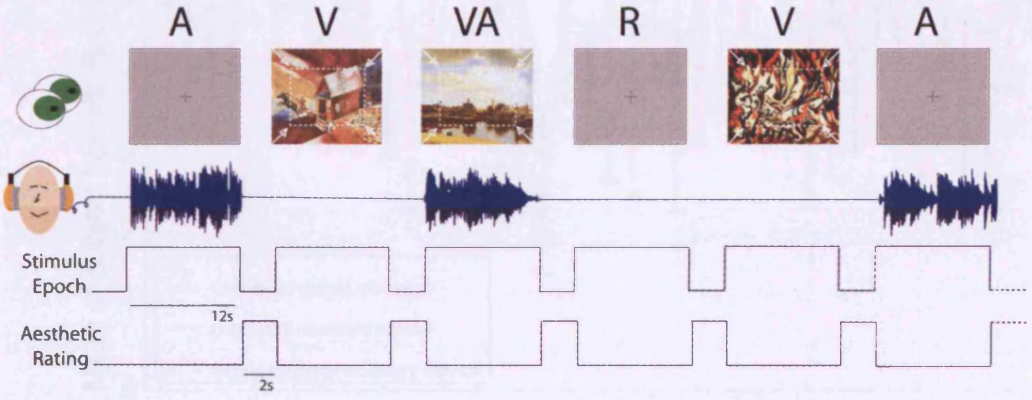


Figure 33: Experiment 4 stimulus presentation paradigm. Subjects were exposed to 12s epochs of visual only (V), auditory only (A), combined visual-auditory (VA) and rest (R) conditions in pseudo-randomised order, followed by a 2s period in which the subject gave an aesthetic rating.

3.8.5 Experiment 4 data analysis

Pre-processing consisted of spatial realignment, normalisation to the standard EPI template and spatial smoothing with a 12mm Full Width Half Maximum (FWHM)

Gaussian kernel as per section 3.1.3. Three separate random effects analyses were performed to investigate: 1) unimodal visual and auditory stimulus processing, 2) crossmodal integration and 3) a parametric regression analysis (Buchel et al., 1996; Price et al., 1992) where each participants' subjective beauty ratings were used as a parametric variable.

For the first and second analyses, data were partitioned at the first level according to whether or not paired stimuli received a higher aesthetic rating than when rated individually. Crossmodal aesthetic enhancement (CAE) refers to pairs of stimuli, which under bimodal conditions were given a higher subjective aesthetic rating than either of the scores given to the same stimuli under unimodal presentation conditions. Comparing the mean aesthetic ratings attributed during scanning to the bimodal stimuli against those given to the same stimuli during unimodal exposure, enables CAE pairings (figure 34: e.g. stimulus pairs #3, 5, 6, 7, 8, 14, 23, 26 etc) to be distinguished from those that were not.

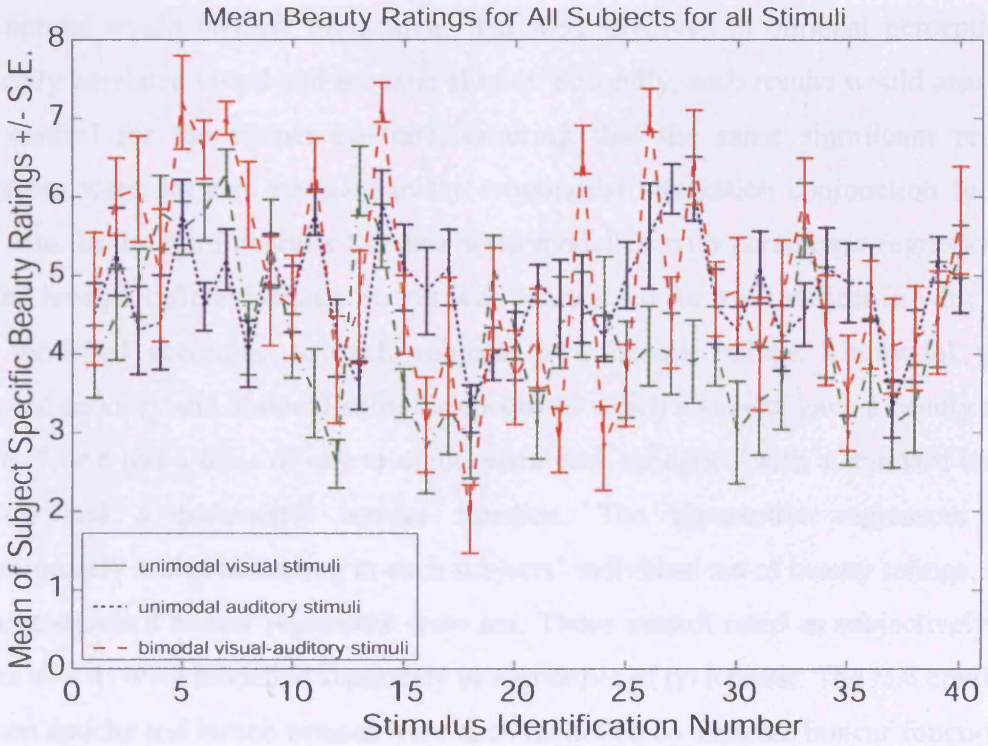


Figure 34: Mean beauty ratings across all scanned subjects. Mean subject ratings (±SE) for the perceived beauty of unimodal visual artworks (green, dot-dash line), unimodal musical extracts (blue, dotted line) and pairs of the same visual artworks & musical extracts presented together (red, dashed line). Several bimodal stimulus pairs were, on average, considered to be crossmodally aesthetically enhanced, i.e. consistently more beautiful when rated in combination than either of the unimodal ratings (e.g. stimulus identification numbers 3, 5, 6, 7, 8, 14, 23, 26… etc.).

The data were fitted with eight separate boxcar functions convolved with the canonical HRF. Three regressors modelled the unimodal visual, unimodal auditory and bimodal conditions for the CAE stimuli, three for the non-CAE stimuli, and one for the rest condition. Response epochs were also modelled as a regressor of no interest.

The first analysis simply consisted of contrasting all unimodal visual and all unimodal auditory epochs, to identify modality-specific brain areas. One CON image per subject was taken from the first level and evaluated at the second level in one-tailed SPM{t} maps. In the second analysis, crossmodal integration (CMI) conjunctions (section 3.1.3, data analysis IV) were carried out separately for CAE and non-CAE stimuli. One-way ANOVAs were used at the second level for the CMI conjunction analyses between the visual-auditory interaction [VA>V]>[A>rest] and bimodal elevation control [VA>V+A]. The aim of the CMI conjunction for the CAE data was to identify the neural correlates of crossmodal enhancement. The crossmodal integration conjunction performed on non-CAE data provided two sources of information. Firstly, this contrast would identify brain areas that were involved in bimodal perception of physically unrelated visual and acoustic stimuli. Secondly, such results would also serve as a control for the former contrast, ensuring that the same significant regional activations were not also revealed in the crossmodal integration conjunction for non-CAE data. In the third analysis the data were modelled with parametric regressors. At the first level, a different design matrix was constructed for each subject in which data were modelled according to each subjects own beauty ratings. Unimodal visual, unimodal auditory and bimodal stimulus epochs in which a subject gave a beauty rating of 5, 6, 7 or 8 (on a scale of one to eight) were each modelled with a standard box car function and a parametric boxcar function. The parametric regressors were proportionately scaled according to each subjects' individual set of beauty ratings, while the three standard boxcar regressors were not. Those stimuli rated as subjectively ugly (scores of 1-4) were modelled separately as regressors of no interest. The rest condition, question epochs and button presses were each modelled by separate boxcar functions as regressors of no interest. All regressors were convolved with the canonical HRF before entering into the design matrix. Voxels at which the standard boxcar regressor accounted for a significant proportion of the variance were revealed in the zeroth order regression, whilst the first order regression revealed areas where the activity was positively correlated with subjective beauty ratings.

3.9 Experiment 4 results: visual-auditory neuroaesthetics

The primary aim of this experiment was to examine brain areas involved in the evaluation of visual and auditory aesthetic beauty. Unimodal visual and auditory epochs were contrasted first to characterise the distribution of unimodal brain activity in response to complex artistic stimuli. The crossmodal integration conjunction was then used to isolate brain regions at which increased activity during bimodal processing of these arbitrarily paired and temporally uncorrelated stimuli induced greater activity than during unimodal stimulus conditions. The parametric regression analysis characterised brain areas at which the magnitude of the BOLD response was positively correlated with each subject's individual set of aesthetic ratings.

3.9.1 Experiment 4 unimodal subtraction analyses

Contrasting unimodal visual versus auditory epochs (V>A) revealed significant voxels throughout the posterior, lateral and ventral occipital cortex (figure 35). It is not possible to confidently distinguish individual visual brain areas using this contrast and was never the aim of such unimodal contrasts (see section 3.1.3). This contrast merely established where greater BOLD activity was detected in response to a broad variety of different visual art pieces than during stimulation with various musical extracts (and *vice versa*). It is nevertheless possible to get a crude idea of the functionally specialised areas involved based on the anatomical distribution of significant activations.

The medial occipital pole activations are likely to correspond to areas V1/V2 (Press et al., 2001) and presumably activation of the superior occipital gyrus (SOG) includes the territory of V3/V3A (Zeki et al., 2003). The strong activation of lateral occipital regions probably reflects the involvement of several specialised visual areas. These are likely to include the human motion area V5 (Zeki et al., 1991), an adjacent area thought to be specialised for processing optic flow (Morrone et al., 2000) and object processing areas of the lateral occipital complex (LOC) (Malach et al., 1995). The stimuli for this experiment used optic flow to create the 'zoom' effect and consisted of multiple complex objects, which together account for these activations. The ventral occipital activation covered the breadth of the fusiform gyri bilaterally extending into the ventral temporal cortex, a region known as the ventral occipitotemporal (VOT) area, which has also been previously noted for its involvement in visual object processing (Grill-Spector and Malach, 2004). This region almost certainly includes the territory in

which the human colour area, the V4 complex, and possibly also the nearby fusiform face area (FFA) (Kanwisher et al., 1997) and/or the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998) reside. Given that the visual art stimuli included a broad variety of highly complex, coloured forms, depicting abstract scenes, landscapes, objects and occasionally animals, bodies or faces, this broad distribution of activity across all levels of the visual hierarchy was in accordance with expectations.

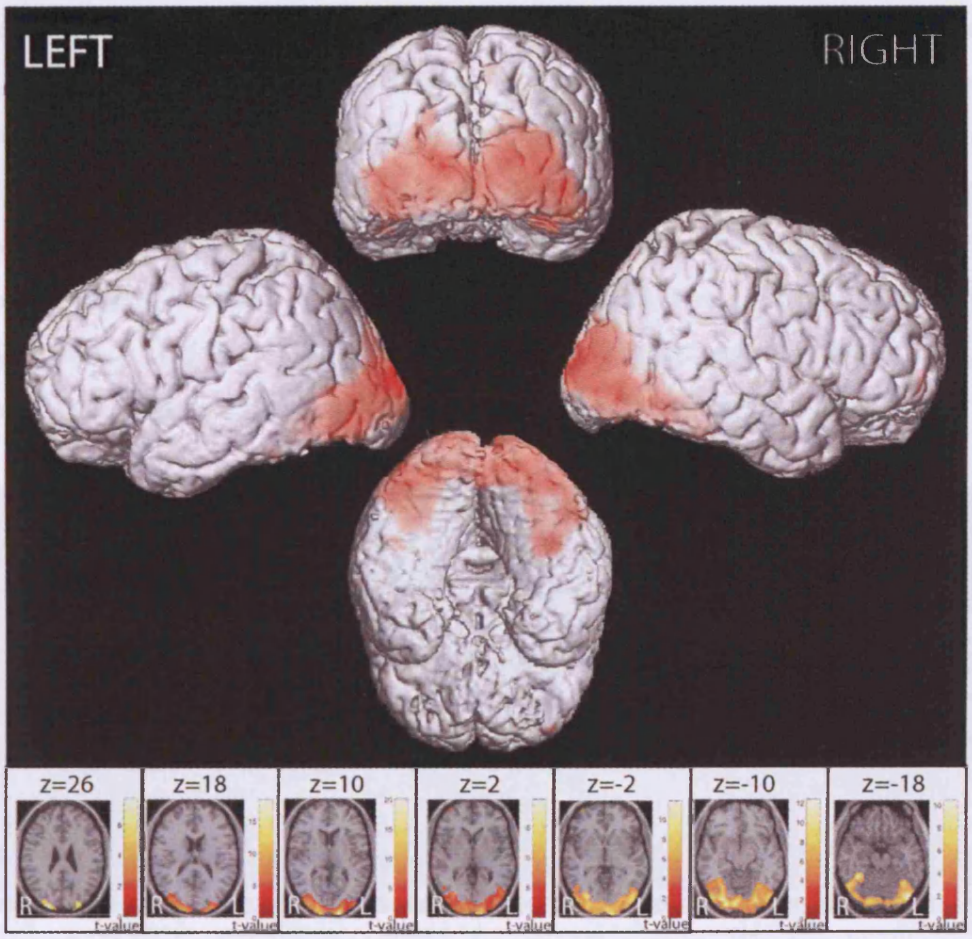


Figure 35: Experiment 4 unimodal visual versus auditory contrast. This revealed significant activity (p<0.001, uncorrected) throughout visual cortex spreading from the medial occipital cortex (V1/V2), dorsally into V3/V3A, laterally into V5 / LOC and ventrally into ventral occipitotemporal cortex. Activations are surface rendered onto a normalised structural image with the cerebellum removed (upper panel) and overlaid on descending structural slices (lower panel).

Contrasting unimodal auditory versus visual epochs (A>V) revealed extensive significant activity (thresholded at p<0.00001, uncorrected) throughout the auditory cortex along the full length of the superior temporal gyri, bilaterally (figure 36). Auditory processing of such complex acoustic features was found to invoke the

participation of the primary auditory cortex located within Heschl's gyrus bilaterally (figure 36, slice z=2) and extensive regions of extra-primary auditory cortex both in anterior and posterior parts of the superior temporal plane / gyrus (figure 36). This is in agreement with various previous auditory imaging studies (Griffiths et al., 1998; Griffiths et al., 2004; Hall, 2002; Patterson et al., 2002; Zatorre and Belin, 2001). This contrast also revealed bilateral activity in the insula and head of the caudate (figure 36, slice z = 2) and in the left cerebellum (figure 36, left-side view). The panel of descending slices (figure 36) illustrates that these data seem to fit well with the proposal that melodic pitch processing invokes activity in cortical areas that progress anterolaterally (z=2 to -14) away from the primary auditory cortex (Patterson et al., 2002). The extensive activation of auditory cortical territory reflects the recruitment of many different areas involved in processing the multitude of complex, spectrotemporal auditory features contained within the musical stimuli. These pieces involve complex arrangements of a variety of musical instruments, which each contribute their unique timbre to the tempo, rhythm and harmony of the overall auditory experience.

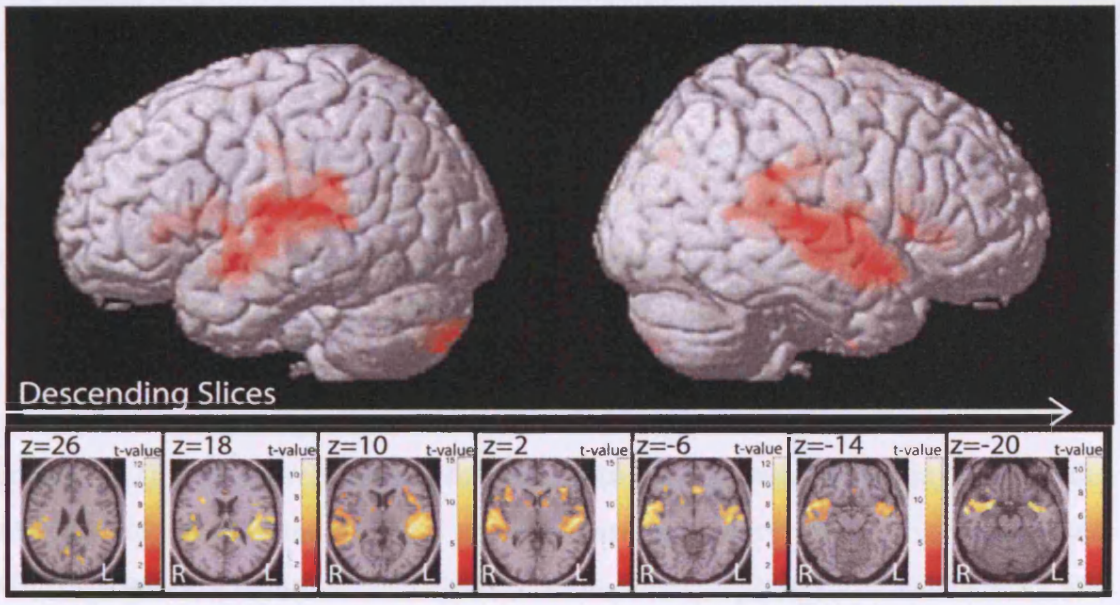
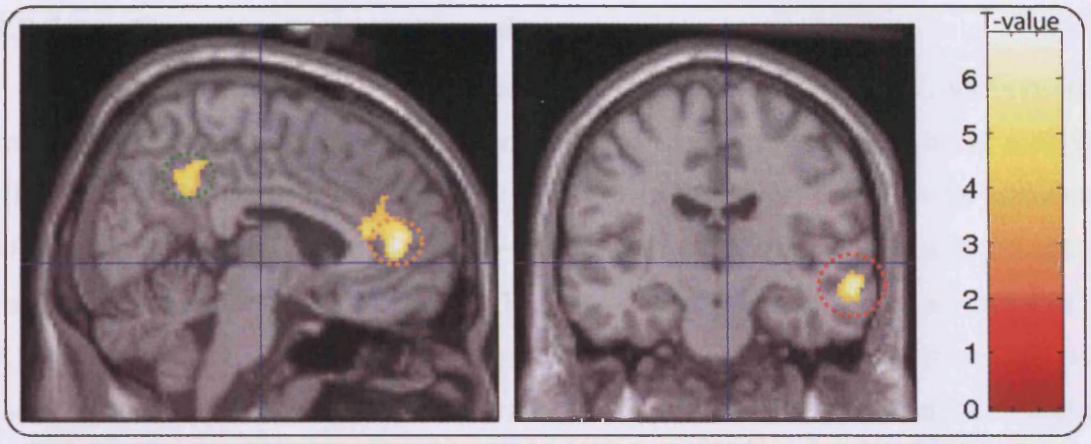


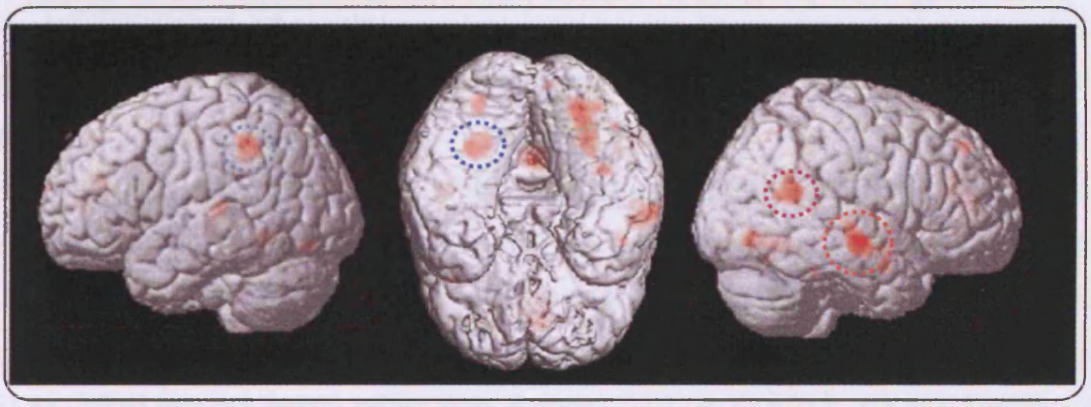
Figure 36: Experiment 4 unimodal auditory versus visual contrast. This revealed significant activity (p<0.00001, uncorrected) throughout auditory cortex surface rendered on the canonical SPM structural image (upper panel) and overlaid on descending slices (lower panel). Activity within parts of bilateral Heschl's gyri are clearly defined in the slice at z=2mm extending caudally into the planum temporale in ascending slices (z=10, z=18, z=26) and rostrally into the planum polare in descending slices (z=-6, z=-14).

3.9.2 Experiment 4 crossmodal integration analysis

The crossmodal integration conjunction was performed separately for crossmodally aesthetically enhanced (CAE) and non-CAE stimulus pairs with the aim of capturing the neural correlates of crossmodal aesthetic enhancement. In fact the results of these separate analyses were virtually identical and so the crossmodal integration conjunction including all data is considered here. The conjunction between the visual-auditory interaction contrast [VA>V]>[A>Rest] and the crossmodal elevation control [VA>V+A] revealed several significant regional brain activations (figure 37). On the lateral surface of the brain these included two loci within the right superior temporal sulcus, with the maximally significant voxel (Z=5.25) midway along the temporal lobe [64,-14,-10, MNI] and a further less significant voxel (Z=3.89) at the posterior end [50, -50, 16, MNI]. The latter activation was located in the immediate vicinity of the pSTS/MTG multimodal area reported in a recent paper by the Beauchamp group (Beauchamp et al., 2004b). On the medial cortical surface there were a pair of cingulate activations, one in the anterior cingulate [6, 46, 8, MNI] and one in the posterior cingulate [0,-46,44, MNI] cortex. Activation of the anterior cingulate has also been reported in another recent study that investigated interactions between congruent versus incongruent bimodal stimuli (Laurienti et al., 2003). On the ventral occipitotemporal surface there was bilateral activation along the collateral sulci [26, -74, -6 and –22, -54, -8, MNI] within a very similar area of visual ventral occipito-temporal cortex revealed in the unimodal visual versus auditory contrast (figure 35). A small voxel cluster was identified in the left supramarginal gyrus [-58,-44,42, MNI].

These findings reflect areas that become more active during bimodal than unimodal conditions, regardless of the fact that intermodal invariances (described in section 1.1.2) were entirely absent. In other words, these areas are not involved in the extraction of information common to both percepts relayed through the visual and auditory sensory systems, as the sources of these stimuli were entirely independent and unrelated. Rather these results may indicate brain regions involved in crossmodal integration of any visual and auditory stimuli considered in parallel giving rise to unified bimodal conscious experience. These regions should be considered as distinct from those involved in integrating visual and auditory component stimuli arising from the same environmental event (discussed further in section 4.3).





Statistics: *p-values adjusted for search volume*

set-level		cluster-level			voxel-level						
p	c	$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	*T-value*	$(Z_=)$	$p_{uncorrected}$	x,y,z {mm}	
0.001	28	0.001	384	0.000	0.012	0.005	8.27	5.25	0.000	64	-14 -10
					0.914	0.037	4.75	3.77	0.000	52	0 -22
					0.999	0.058	4.04	3.37	0.000	48	-2 -14
		0.000	665	0.000	0.093	0.015	6.80	4.73	0.000	6	46 8
					0.362	0.022	5.82	4.31	0.000	-4	38 22
					0.561	0.026	5.43	4.13	0.000	4	32 18
		0.001	404	0.000	0.110	0.015	6.69	4.68	0.000	26	-74 -6
					0.867	0.035	4.87	3.84	0.000	30	-62 -10
					0.998	0.053	4.17	3.44	0.000	36	-76 -4
		0.002	365	0.000	0.113	0.015	6.67	4.67	0.000	0	-46 44
		0.269	100	0.023	0.486	0.024	5.57	4.19	0.000	-58	-44 42
		0.059	171	0.004	0.585	0.026	5.39	4.11	0.000	-22	-54 -8
					0.641	0.027	5.30	4.06	0.000	-30	-52 -10
		0.163	123	0.013	0.715	0.030	5.17	4.00	0.000	-42	-30 8
		0.017	235	0.001	0.822	0.033	4.97	3.89	0.000	50	-50 16

Figure 37: Experiment 4 crossmodal integration conjunction. This CMI conjunction reveals areas at which visual stimulus processing was modulated by concurrent auditory stimulus processing and greater responses were detected under bimodal versus unimodal stimulus conditions. Activations were overlaid on saggital and coronal sections of the canonical SPM structural image (upper panel), surface rendered on the canonical SPM brain (middle panel, left and right views) and on a normalised structural image with the cerebellum removed to reveal ventral occipitotemporal surface (middle panel, ventral view). The results table is also provided (bottom panel) with MNI coordinates colour-coded to match the corresponding dashed circles.

3.9.3 Experiment 4 zeroth order parametric regression

Voxels at which the BOLD response was significantly ($p<0.001$, uncorrected) correlated with the zeroth order regressor (unscaled boxcar function convolved with the HRF) for visual, auditory and bimodal epochs were surface rendered onto the canonical SPM structural brain (figure 38). Figure 38 consists of a colour-coded map of voxels rendered on the surface the canonical SPM brain, illustrating unimodal activations (V, A), bimodal activations (X) and also regions of overlap between unimodal and bimodal (VX, AX) and between all three conditions (VAX). This reflects areas activated during sensory processing and cognitive evaluation of beauty, irrespective of the magnitude of subjective beauty ratings. Beauty evaluation during exposure to visual artworks resulted in active voxels in bilateral posterior occipital [peak activations: ±22,102,18, MNI], right ventral occipital [36,-64,-12, MNI], left intraparietal sulcus [-32,-64,58, MNI] and left lateral orbito-frontal cortex [-42,54,-6, MNI]. Voxels selectively responsive during aesthetic evaluation of auditory stimuli were identified in bilateral superior temporal gyrus [peak activations at: -50,-18,4 & 52,6,-37, MNI], left intraparietal sulcus [-32,-62,58, MNI] and left lateral orbito-frontal cortex [-42,54,-6, MNI]. During exposure to bimodal stimulus pairs, significant voxels were identified ($p<0.001$, uncorrected) in all of the above regions, i.e. bilateral occipital [±22,102,18, MNI], bilateral superior temporal gyrus [64,-14,2 & -50,-18,4, MNI], left lateral orbito-frontal [-48,50,-6, MNI] and left superior parietal cortex [-32,-62,58, MNI].

A conjunction analysis was performed to formally identify areas involved in the beauty evaluation irrespective of whether the information was visually or acoustically perceived (figure 39). This formally confirmed the presence of significant ($p<0.05^2$, corrected for multiple comparisons) activity profiles at voxels within the left inferior parietal sulcus [-32,-62,58, MNI; k=125; z=5.28] and left lateral orbito-frontal cortex [-42,52,-8, MNI; k=118; z=4.43] at the same sites observed to be regions of overlap in figure 38. As these areas were active during the evaluation of beauty irrespective of whether the stimuli in question were visual or auditory, they must therefore be involved in supramodal or amodal aspects of the given task. Furthermore, the activity in these areas was not significantly modulated by the degree of beauty in the visual and auditory stimuli as these are accounted for by the first order regressor.

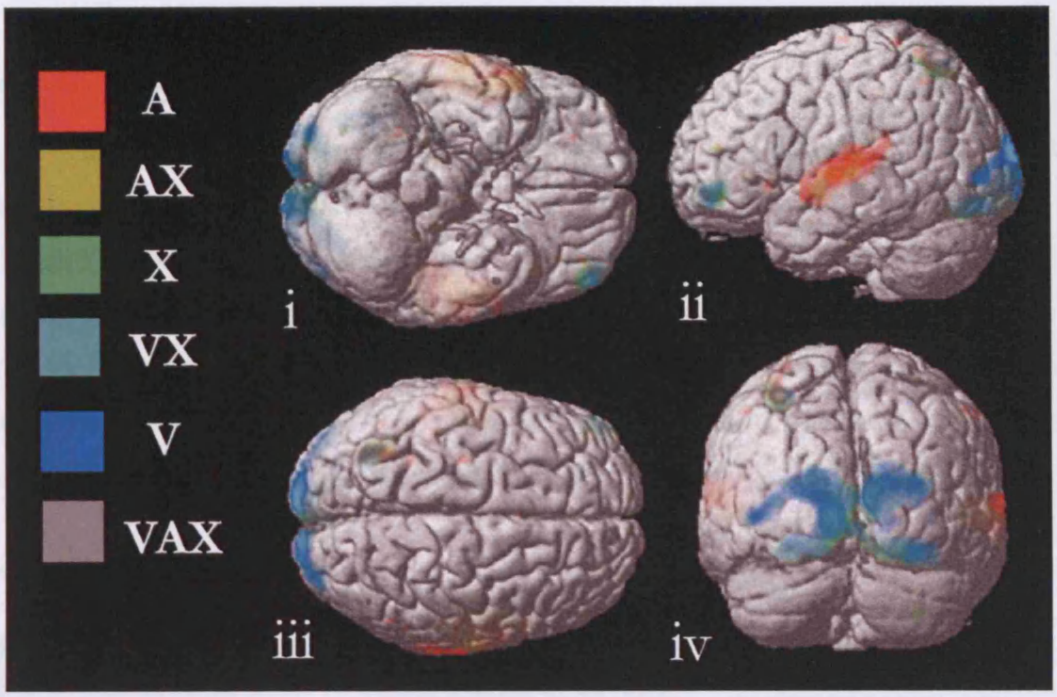


Figure 38: Unmodulated activations during visual, auditory and bimodal beauty evaluations. Voxels responsive to unimodal auditory (A, red), unimodal visual (V, blue) and bimodal (X, green) stimuli irrespective of the beauty rating have been surface rendered on the ventral (i), left lateral (ii), dorsal (iii) and caudal (iv) surfaces of the canonical SPM structural brain image. Regions of overlap between bimodal and unimodal auditory activations are gold (AX), between unimodal visual and bimodal activations are cyan (VX) and between all three are coloured grey (VAX).

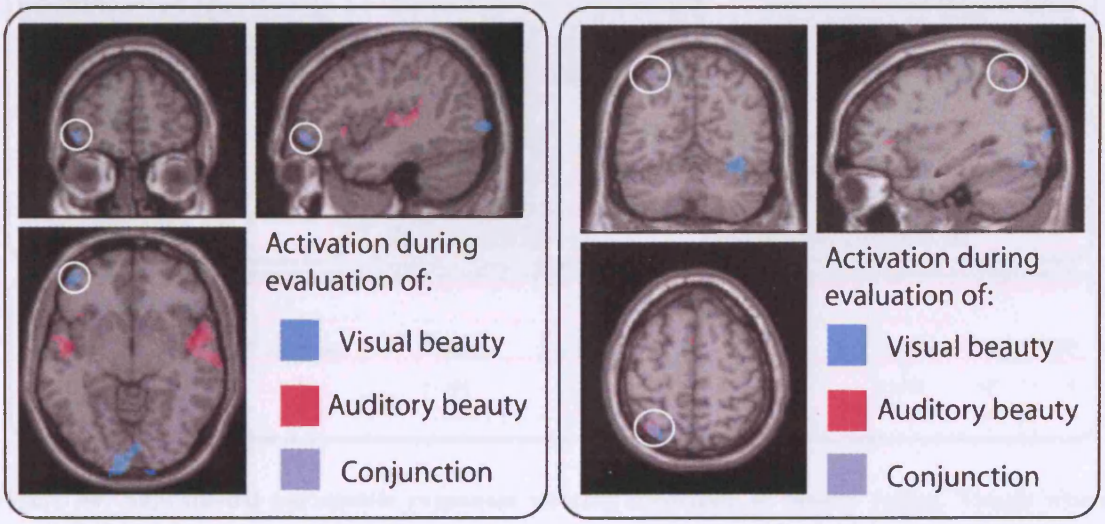
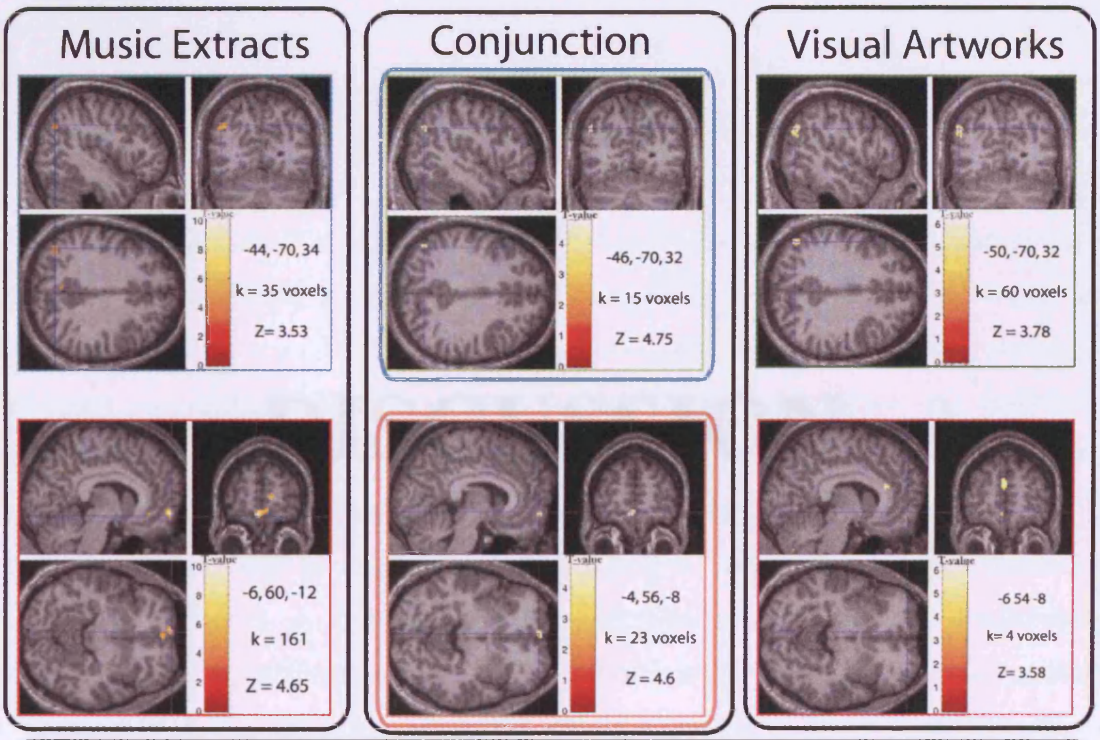


Figure 39: A conjunction analysis between unmodulated responses during visual and auditory beauty evaluation. This confirmed that significant ($p<0.05^2$, corrected) activation in the left lateral orbitofrontal cortex (left panel) and the left intraparietal sulcus (right panel) corresponded to the evaluation of beauty in visual and auditory artworks irrespective of the beauty rating. The activations have been overlaid on the coregistered and normalised structural brain image of a representative subject (WM).

3.9.4 Experiment 4 first order parametric regression

The first order parametric regression analysis revealed voxels at which BOLD activity was positively correlated with the subjective beauty rating attributed to unimodal visual art pieces or unimodal musical extracts (figure 40). A conjunction analysis ($p<0.001^2$, uncorrected) was performed to confirm that BOLD responses at the left parieto-occipital junction (top row) and medial frontal cortex (bottom row) were parametrically modulated according to the magnitude of beauty rating attributed to both unimodally presented musical extracts (left column) and artworks (right column).



cluster-level			voxel-level					
$p_{corrected}$	k_E	$p_{uncorrected}$	$p_{FWE-corr}$	$p_{FDR-corr}$	T	(Z_e)	$p_{uncorrected}$	x,y,z {mm}
	15		0.001	0.001	4.75	5.69	0.000	-46 -70 32
	23		0.002	0.001	4.60	5.57	0.000	-4 56 -8

Figure 40: Supramodal parametric responses varying according to beauty rating. Voxels where activity was positively correlated with subjective aesthetic beauty ratings were identified at the left parieto-occipital junction and the medial frontal pole. Significant ($p<0.001$, uncorrected) parametric modulation of BOLD responses to musical extracts (left column), to visual artworks (right column) and significant ($p<0.001^2$, uncorrected) conjunction of these two parametric regressions (middle column) were overlaid on slices of the normalised structural image. The table of results for the conjunction analysis has been colour-coded to match the corresponding boxes.

Part 4:

DISCUSSION

4.1 Discussion overview

The experiments comprising this thesis have approached the question of visual-auditory perceptual integration from several different perspectives. These have revealed neuroanatomical sites at which visual and auditory stimuli are processed during evaluation for object recognition (experiment 1), the detection of crossmodal (experiment 2) and/or intramodal (experiment 3) synchronicity, or in light of a common subjective metric (i.e. beauty, experiment 4).

In experiment 1, all bimodal stimuli were presented synchronously to enable visual-auditory integration to result in unified perception. When subjects identified degraded visual and auditory video footage of everyday occurrences under both bimodal and unimodal conditions, activity profiles consistent with crossmodal integration were revealed at the posterior temporo-parietal junction (TPJ) only. This observation indicates that visual and auditory parts of individual environmental phenomena may be perceptually integrated at the posterior TPJ.

In experiments 2 and 3, the synchrony between abstract, visual and auditory stimuli was specifically manipulated to investigate where in the brain crossmodal temporal covariation might be monitored. Both experiments revealed significantly increased right-lateralised activity at the anterior frontal operculum and underlying anterior short insula gyrus, under crossmodally asynchronous versus synchronous conditions. As discussed in section 1.2, any neural mechanism responsible for regulating multisensory grouping would be expected to exhibit sensitivity to crossmodal synchronicity, i.e. to discriminate between temporally correlated and uncorrelated bimodal stimuli. Therefore, sensitivity of this brain region to crossmodal asynchrony, suggests that it may be involved in preventing crossmodal integration of unrelated bimodal stimuli into unified perception. In experiment 3 only, the amygdala was found to be more active, bilaterally, during crossmodal synchrony versus asynchrony. In this experiment two visual and two auditory stimuli were presented in all conditions, whilst in experiment 2 only one visual and one auditory stimulus was presented. The elevated amygdalar activity may therefore reflect the increased salience of crossmodally covarying stimuli embedded in a multiple stimulus array. It is also possible that it is actually responsible for the perceptual pairing of a sound to the appropriate visual object, under conditions where several combinations are potentially viable.

A region of medial frontal cortex at which response magnitude was positively correlated with subjective beauty ratings was identified in experiment 4. This brain area was parametrically modulated according to the degree of perceived beauty, irrespective of whether the target of the aesthetic evaluation was received via visual or auditory sensory apparatus. Hence this observation reveals the neural correlate of supramodal aesthetic contemplation at a cerebral site far removed from both visual and auditory sensory cortices.

4.1.1 Constructing the argument

The historical survey was designed to provide a theoretical background in light of which the reasoning and motivation that guided the approach to these studies of visual-auditory integration might be fully understood. Having established the history of functional specialisation in the visual brain and the recent observations in the auditory cortex that are more or less compatible with this (section 2.1), the psychophysical evidence establishing the existence of performance improvements in human perception that results from merging visual and auditory sensory information was outlined (section 2.2). Neuroanatomical sites at which the visual and auditory information might become integrated to support such behaviour were then proposed, based on the findings of anatomical tracer studies and electrophysiological investigation in the primate (section 2.3), human neurological studies (section 2.4) and non-invasive human imaging studies (section 2.5).

The rest of this section highlights additional background information to provide a stable platform upon which the key findings outlined above can be discussed in greater depth.

4.1.2 Terminological definitions

Different research groups use the various terminologies that describe experiments investigating visual and auditory sensory processing in different ways, which can lead to a certain degree of confusion (Calvert, 2001). Electrophysiologists were the first to coin the term 'multisensory integration' in the context of the behaviour of single neurons (Stein, 1993). In the functional imaging literature, although some have directly adopted this term to describe functional brain activations involved during exposure to stimuli in more than one sensory modality (Calvert, 2001), many have

introduced new terms to describe their findings e.g. 'polymodal processing' (Bremmer et al., 2001), 'multimodal sensory integration' (Barraclough et al., 2005), or 'multisensory facilitation' (Schnupp et al., 2005). There are many inconsistencies in the use of other commonly used terms like 'multimodal', which can describe not only brain areas sensitive to stimulation in more than one sensory modality, but also the use of several different methodological approaches to a certain problem, e.g. using fMRI, TMS and EEG. In addition, whereas some authors use certain terms interchangeably, others have stricter definitions for the use of specific terms. For absolute clarity, the specific definitions of the terminology used in this discussion will be outlined.

Bimodal / Unimodal

The majority of electrophysiological studies investigating neuronal responses to visual and auditory stimuli were limited to the detection of unimodal responses to each stimulus, without investigating responses to simultaneous stimulation in both sensory modalities (see section 2.3.3). Consequently, in electrophysiology, neurons are described as bimodal if they respond to both unimodal visual and unimodal auditory stimulation. However, in the discussion of functional imaging results, the term 'bimodal' is usually used to refer to *concurrent,* visual-auditory stimulation and the regional brain responses that result from such stimulation.

Bimodal objects

In this thesis the term 'object' is used to describe sensory phenomena that can be characterised visually, acoustically or by the combination of visual and acoustic features. In the context of experiment 1, the objects in question were a selection of events and actions performed by human actors. It may seem contrary to convention to use the term 'object', which is generally used to refer to static, independent, permanent, visual items (such as a ball, rope or cup, for instance), rather than actions performed with these items. However, the dictionary definition in fact allows for its use in this context, describing an 'object' as: 'something perceptible by one or more of the senses' (www.dictionary.com). This does not appear to prohibit its deployment in descriptions of dynamic sensory events, particularly as auditory objects are by their very nature dynamic and impermanent. Thus, the bimodal badminton 'object', despite referring to groups of events: the racquet hitting the shuttlecock and the players' footfalls / arm movements that actually cause the visual-auditory components, can nevertheless be

described in terms of concurrent visual and auditory percepts that together characterise the multisensory percept of badminton. This is not stated in an attempt to incite a revolution in the way the word 'object' should be used in science and literature, but simply to facilitate the discussion of the results by using a single term to describe visual and auditory sensory cues that pertain to the same environmental phenomenon, irrespective of the agent(s) that caused the sensory cues to arise.

Multisensory convergence

The term, 'multisensory convergence' (MSC), is used to describe voxels at which significant responses are detected during concurrent visual and auditory stimulation. The word 'convergence' conveys the meaning that visual and auditory stimuli induce activity at the same location (i.e. voxel), akin to Meredith's areal convergence (Meredith, 2002). Electrophysiology can distinguish between convergence of visual and auditory responses at single or adjacent neurons. Functional MRI, on the other hand, does not have sufficient spatial resolution to make distinctions at the neuronal level and so bimodal activations could reflect integrated, crossmodal stimulus processing, but could equally reflect intermingled unimodal neuronal populations within a single voxel. This is because in fMRI the BOLD activity at each voxel reflects the averaged metabolic demands of several thousand neurons and so simple subtraction analyses cannot distinguish whether unimodal or bimodal neuronal activity induces such activation profiles. Unless crossmodal integration has been specifically demonstrated by appropriate statistical testing, bimodal responses can only be attributed to MSC.

Crossmodal integration (CMI)

The use of the term 'CMI' is reserved for contrasts that specifically identify voxels at which activation profiles indicate significant interactions between the processing of visual and auditory stimuli when presented concurrently, that is under bimodal conditions (VA), as compared to separately, i.e. unimodal visual (V) and unimodal auditory (A) stimulation. The CMI conjunction (see section 3.1.3-data analysis IV) makes it possible to demonstrate that visual stimulus processing is modulated by the presence of concurrent auditory stimulation and *vice versa*. Therefore CMI, rather than MSC, can be used to describe the response at such voxels as it provides evidence of positive interactions between visual and auditory stimulus

processing. The choice of the word 'crossmodal', over the more commonly used 'multisensory', was made to emphasise the notion of intersection (X) between *two* sources of sensory information at which such visual-auditory interactions occur.

Supramodal

Certain brain areas can be responsive to the presence of specific information within different sensory stimuli, irrespective of the sensory modalities through which it is perceived, e.g. to words whether read or heard (Cohen and Dehaene, 2004). Such modality invariant responses are sometimes referred to as 'heteromodal', reflecting the possibility that these responses may be derived from a mixture of juxtaposed unimodal subregions. The term 'supramodal' is used to describe regional brain activations that appear to be sensitive to features above (supramodal) the level of unimodal sensory processing. The term 'amodal' on the other hand can be used to describe activations that are thought to be unrelated to sensory processing.

In the next two sections the basic principles of visual-auditory bimodal perception will be briefly revisited to assist the discussion of results.

4.1.3 Why integrate bimodal sensory information?

The visual and auditory senses convey information that can be divided into two separable components: the mutually exclusive components such as colour or timbre and the features that can be common to both, such as amplitude, rate, rhythm and spatiotemporal coincidence. Where common cause can be established between visual and auditory stimuli, the possibility of crossmodal integration arises. The confidence with which such causal relationships can be established between visual and auditory stimuli encountered in the natural environment, varies considerably according to how easily the visual and auditory changes can be resolved. At one extreme there are the qualitative crossmodal correlations, such as those that exist between the sight and sound of trees rustling in the breeze. At the other extreme are the tightly phase-locked, quantitative crossmodal phenomena, such as the sight/sound of a musician playing the drums.

One way in which man has adapted to the environment is by setting up perceptual predictions about how a sensory event will unfold. The reliability of these predictions varies depending on how well the stimulus has been characterised up to that

point in time. If the sensory event comprises both an acoustically and a visually detectable sequence of changes, with a distinct pattern that is identifiable and can be confirmed through repetition, then predictions can be set up for the future course of the two independently perceived channels of information. The independence originates from the fact that completely different stimulus energies are transduced into neuronal action potentials and then processed separately in distinct regions of sensory cortex, ensuring the fidelity of the resulting unimodal percepts. When the two signals happen to be correlated in some way, i.e. having originated from the same environmental event, each signal can mutually reinforce the other. Put another way, a repeating pattern in the visual stimulus predicts the future course of not only the visual, but also the acoustic stimulus and the pattern in the acoustic stimulus can also predict not only the future course of the acoustic stimulus, but also the correlated visual stimulus. Having established covariation between a pair of environmental, audiovisual, sensory stimuli, any redundancy in the signal could be integrated across visual and auditory domains for bimodal perception. From an evolutionary point of view, the perceptual improvements resulting from integration of appropriate bimodal sensory information may well have conveyed considerable survival advantages, for instance by enabling faster and more accurate distinction between prey and predator.

4.1.4 The regulation and realisation of crossmodal integration

Given the wide separation of cortical territories involved in processing visual and auditory sensory input, it is not clear how appropriate information is integrated during concurrent bimodal stimulation. The neural correlates of crossmodal integration can be broken down into two parts: (1) where in the brain are visual and auditory stimuli brought together in unified conscious perception, and ultimately, (2) which brain regions are involved in the regulation of whether bimodal stimuli are considered single or separate entities. The debate regarding the mechanism by which perceptual integration is achieved in the brain can be separated into two main groups. On the one side some researchers, inspired by retrograde tracing studies in monkeys identifying direct connections between primary auditory cortex and visual brain areas (Falchier et al., 2002), champion the role of enhanced intermodal communication (e.g. Schroeder et al., 2003). In the other camp, there are those who have identified areas of multisensory

cortex at which visual and auditory stimuli appear to become integrated, supporting the crossmodal hypothesis (e.g. Beauchamp et al., 2004b).

There is evidence to support the claims of both groups. In support of the direct intermodal influence lobby, a few studies have demonstrated 'unimodal' visual cortex activation in response to auditory stimulation (Zimmer et al., 2004). Others have identified 'unimodal' auditory cortex activation in response to visual stimulation (Calvert et al., 1999). This has led to considerable excitement, but these findings seem to consist of a few isolated examples rather than a comprehensive body of evidence. However, the support for the opposing view of multisensory integration invoking convergence and integration at dedicated CMI areas is equally sparse. Early electrophysiological studies made robust demonstrations of significant responses to both visual and auditory inputs at the superior temporal polysensory area (Baylis et al., 1987; Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988), which probably corresponds to the site of multisensory integration proposed in the Beauchamp studies (Beauchamp et al., 2004a). In addition, neurons responsive to both visual and auditory stimulation were also demonstrated in primate prefrontal cortex and in particular, the orbitofrontal cortex (Benevento et al., 1977; Ito, 1982; Petrides et al., 1978).

The important variables involved in discriminating between those multisensory cues that should be crossmodally integrated and those that should not, have largely been settled in the psychophysical literature (section 2.2). Spatial and temporal coincidence is vitally important to this process. Of these, it seems that when the bimodal stimuli are temporally complex, with both visual and auditory components evolving gradually over time, temporal crossmodal covariation appears to be the dominant force in multisensory grouping. For instance, in the ventriloquist illusion the spatial disparity between the source of sight and sound is discounted in favour of the tight synchrony between the dummy's lips and the performer's voice. In the functional imaging literature, the neural correlates of visual and auditory spatial or temporal correspondence have not been thoroughly pursued, with only a handful of studies investigating these issues. Whilst several studies have investigated the neural correlates of congruent and incongruent visual and auditory speech cues (e.g. Macaluso et al., 2004), only two previous imaging studies have investigated the neural correlates of synchronicity for non-speech stimuli (Bushara et al., 2001; Calvert et al., 2001).

4.1.5 The strategy for investigating visual-auditory integration

The findings of the visual-auditory identification, synchrony and beauty studies are all discussed with the primary aim of attempting to distinguish between sites of unimodal and bimodal/crossmodal processing. Care has been taken to highlight findings that confirm, contradict and extend prior knowledge pertaining to mechanisms of visual-auditory integration and its regulation, as they are currently understood.

In order to pursue the question of where in the brain visual-auditory integration occurs, various fMRI paradigms were used in which the BOLD response under bimodal and unimodal conditions could be compared in the context of a resting or perceptually neutral baseline. This approach was taken in an attempt to identify previously undiscovered sites of crossmodal integration (CMI). Whilst congruent visual and auditory stimuli arising from the same environmental events were used in experiment 1, in the other experimental paradigms conducive to the CMI conjunction analysis (experiments 2 and 4), visual and auditory stimuli were unrelated. CMI activation profiles identified in experiments 2 and 4 would thus reveal brain areas involved in processing concurrently perceived, but unrelated stimulus pairs. The aim was to characterise the brain areas involved in such non-specific CMI activation in order to elucidate the true neural correlates of visual-auditory integration of bimodally detected sensory stimuli arising from the same event (i.e. experiment 1).

As discussed earlier, CMI could, in theory, be achieved either by increasing inter-communication between unimodal sensory cortices (directly or indirectly via corticothalamocortical or cortico-cortical loops), or through convergence at separate, dedicated, crossmodal areas. In order to establish whether CMI might be achieved by amplification of intercommunication between 'unimodal' cortices, unimodal contrasts were performed to define the visual and auditory brain areas involved in each experimental paradigm. The unimodal regional brain activations from the three experiments in which unimodal contrasts were possible are discussed in the next section (section 4.2). It was then possible to demonstrate whether regional activations detected in the CMI conjunctions were located within or beyond unimodal sensory territories.

By comparing results of the CMI conjunctions in experiment 1, 2 and 4, it was also possible to investigate whether CMI occurs at distinct sites according to the way in which the sensory information is being evaluated (i.e. for object recognition, synchronicity judgements or for aesthetic evaluation), or whether it occurs at a

dedicated CMI site regardless of the cognitive constraints. These results are discussed in section 4.3.

The crossmodal synchronicity studies of experiments 2 and 3, served to address the second aim of this thesis - to identify brain areas that might be involved the regulation of CMI, i.e. multisensory grouping. In both experiments, simple, abstract, cyclically varying, visual and auditory stimuli were created in order to manipulate synchronicity and identify brain areas that demonstrated a preference for either synchronous or asynchronous stimuli. These experiments comprise the core of this thesis, as the paradigm yielding the key result of experiment 2 was revised in experiment 3, to both confirm and extend the original finding. These results are discussed in section 4.4. In the penultimate section (section 4.6), the key findings from experiments 2 and 3 are drawn together to form a neuroanatomical model detailing the essential nodes through which CMI is regulated.

In experiment 4, entirely unrelated visual and auditory stimuli were evaluated in a single context, that of an aesthetic judgment of beauty. This enabled us to identify parametrically varying beauty responses at multisensory convergent supramodal brain loci. These findings are discussed in section 4.5.

4.2 Characterising unimodal sensory brain areas

In experiment 3, no unimodal epochs were included, precluding comparison of regional brain activity during unimodal visual versus auditory stimulation. However in the other three experiments, visual and auditory brain areas could be revealed using appropriate unimodal contrasts, i.e. [V>A] and [A>V] (results sections 3.3.1, 3.5.1, 3.9.1). Different combinations of specialised visual areas in occipital, ventral temporal and posterior parietal cortex were detected in the unimodal visual contrast [V>A] according to the specific visual features present in each experiment. Equally, activations within the superior / middle temporal and inferior frontal gyri were detected by the unimodal auditory contrast [A>V], the precise distribution of which also varied according to the specific combination of auditory features involved in each experiment. Regions of unimodal sensory cortex found to be involved in processing the visual or auditory stimuli in one experiment, but not the others, could usually be attributed to the presence of different the stimulus features and/or task requirements used in each experiment. For instance, a greater emphasis is placed on extracting dynamic features from noisy, complex, naturalistic stimuli in experiment 1, on rhythm within the simple, abstract, unimodal stimulus cycles of experiment 2 and on the aesthetic value of the relatively complex and highly-detailed aesthetic stimuli of experiment 4.

The distribution of significant voxels across visual and auditory sensory cortices in unimodal stimulus contrasts varies, not only according to the specific features included in the stimulus array and the task requirements, but also according to experiment-specific thresholding effects. Therefore comparison of the spread of unimodal activity between different experiments is of limited use. The unimodal contrasts were primarily included to ensure regional brain activation in the appropriate cortical areas was revealed, thus verifying the data had been properly modelled. However, it is still worth briefly describing commonalities and differences in the distribution of unimodal cortical activations in the context of the specific stimulus features presented in each experiment. This would enable a better informed interpretation of the CMI conjunction results should the same brain loci have been revealed.

4.2.1 Unimodal visual activations

Visual areas commonly identified in more than one experiment can be accounted for by similarities in the visual features present in the visual stimulus arrays used in each experiment. For instance, significant voxels within the anatomical coordinates at which the human motion area V5 is located were identified in all three experiments. This can be attributed to the 'object' motion in experiment 1, to the motion cycles in experiment 2 and to the 'zoom effect' in the visual stimuli of experiment 4. However, the distribution of significant voxels in other unimodal visual brain areas varied between experiments. In the absence of explicit localiser contrasts to distinguish between the different functionally specialised subsections of each particular cortical territory, it is impossible to determine which specific class of visual features contribute to each of these activations. However it is possible to speculate about which visual features present in one experiment, but not the others, are most likely to account for the activation differences.

Unimodal visual activation of early visual areas

The visual stimuli in all three experiments were centrally presented and would therefore have been expected to induce activity at the foveal confluence of early visual areas (V1/V2) located at the occipital poles. Indeed, the V>A contrast revealed activity at the appropriate locus in experiments 2 and 4, but in not in experiment 1. This can be explained by the fact that, in experiment 1 only, visual white noise was presented with all unimodal auditory stimuli as a control measure to validate the CMI conjunction. The absence of significant activation at the foveal confluence in experiment 1 suggests that the inclusion of pure visual white noise during the unimodal auditory conditions controlled well for low-level visual features. Therefore, the visual versus auditory contrast in experiment 1 reveals only visual brain areas involved in the extraction of intermediate to high-level stimulus features from the background noise.

Unimodal visual activation of the lateral occipital cortex

In primate neurophysiology, visual object feature processing invariably invokes activations in the so-called 'what' stream (Ungerleider and Mishkin, 1982), a blanket term referring to areas of extrastriate cortex located throughout ventral and lateral areas of occipital and inferior temporal cortex, particularly implicating areas TE, TEO and IT

in the monkey (Desimone et al., 1984; Gallant et al., 1993; Gross et al., 1972; Logothetis and Sheinberg, 1996; Pasupathy and Connor, 1999; Tanaka, 1996; Vogels, 1999). In a similar vein, the LOC is an umbrella term used to refer to visual object-sensitive areas in man, located both within the lateral occipital cortex and also, confusingly, activations in the ventrally adjacent posterior fusiform gyrus (Kourtzi and Kanwisher, 2000). The LOC is consistently more active during exposure to intact versus scrambled visual objects (Malach et al., 1995; Kanwisher et al., 1996). This is the case regardless of whether the objects are defined by line drawings, shading, texture, motion or luminance and independently of changes in object size or position (Kourtzi and Kanwisher, 2000). It also demonstrates response adaptation upon repeated presentation of the same object even when is has been subsequently defined by a different set of cues (Grill-Spector et al., 1999; Grill-Spector et al., 1998). The lateral occipital part of the LOC is posteroventrally adjacent to V5 (Malach et al., 1995; Self and Zeki, 2004).

In experiment 2, as the visual stimuli consisted of random dot arrays, in the absence of any discernable shapes or objects, no significant visual activity was identified in the vicinity of LOC (apart from V5). In experiments 1 and 4 however, there were many visual shapes and forms to be processed in the film footage and art stimuli, respectively. Consequently, extensive activation of the LOC was revealed by the unimodal visual contrast [V>A], reflecting the high object processing demands in both these experiments. In addition, both experiments also demonstrated visual activations posterodorsal to the LOC, probably indicating the involvement of form processing in area V3/V3A (Zeki et al., 2003).

The activation of LOC appeared to be differently distributed in experiments 1 and 4. There are several possible causes for this, of which the marked stimulus differences between the two experiments are likely to make a significant contribution. There was great variation in the visual stimuli of experiment 4, but one feature that was ever-present was the zoom effect, which did not feature in experiment 1. This may therefore account, in part, for the different distribution of activity between the two experiments in the vicinity of the LOC, as the region of lateral occipital cortex thought to be specialised for processing optic flow (Howard et al., 1996; Morrone et al., 2000) would be implicated in one, but not the other experiment. Furthermore, the visual stimuli of experiment 1 often involved moving human body parts and so greater activation was expected in other subregions of the LOC in this experiment, due to the involvement of other specialised visual regions. These include the extrastriate body area

(EBA) (Downing et al., 2001) and another region thought to be specifically responsive to biological motion in the right posterior STS, just anterior to V5, during recognition of these visual 'objects' (Grossman et al., 2000; Grossman and Blake, 2002).

Unimodal visual ventral occipito-temporal activations

The involvement of ventral occipitotemporal (VOT) cortex in visual object processing is well established (Grill-Spector and Malach, 2004). Subdomains of the VOT have been discerned in previous imaging experiments indicating regions that are particularly responsive to faces, i.e. the fusiform face area (FFA) (Kanwisher et al., 1997) and places, i.e. the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998). Furthermore, rostral parts of the VOT have been implicated specifically during visual object recognition (Bar and Biederman, 1999; Bar et al., 2001).

In the absence of shapes and forms in experiment 2, the unimodal visual versus auditory contrast gave rise to minimal activation of the VOT. However, due to the plentiful supply of shapes and forms in the film clips of experiment 1 and the various art stimuli of experiment 4, extensive bilateral fusiform activity was observed in both experiments. This bilateral fusiform activity extended rostrally from lateral parts of the posterior fusiform (ventral LOC) into ventral temporal regions of the fusiform gyrus adjacent to the collateral sulcus. Posterior fusiform gyrus activations are often found to overlap with the rostrally adjacent ventral occipito-temporal cortex (VOT) (Grill-Spector and Malach, 2004).

Active voxels extended rostrally across the collateral sulcus into parahippocampal territory, bilaterally in experiment 1 and on the right side only in experiment 4. This raises the possibility that the PPA is activated due to processing of the settings in which the films were shot in experiment 1 and of the places depicted in several of the art stimuli of experiment 4. Posterior to this and about midway along the fusiform gyrus lies both the V4 complex and the FFA. Most of the art stimuli in experiment 4 were coloured, accounting for this activation in part by the involvement of the V4 complex in colour processing. Although the visual stimuli in experiment 1 were all black and white, human faces were often visible and so the participation of the FFA may account for some of the VOT activity in this experiment.

Unimodal visual activations at the parietal-occipital junction

The superior occipital gyrus was activated in all three experiments, probably reflecting motion related activity in area V3A/V7 (Smith et al., 1998), but in experiment 1 only, significant voxels crossed the parietal-occipital junction and encroached into posterior parietal cortex. These activations could be interpreted as visual areas that support the extraction of two- and three-dimensional visual features from the very low resolution and low salience moving images used in this experiment. Indeed, a whole branch of visual imaging studies investigating motion feature processing have implicated area V3A in 2-D shape extraction and 3D structure from motion (Orban et al., 1999; Vanduffel et al., 2002). The parietal visual activations in experiment 1 appear to consist of two separate clusters which may correspond to the ventral/parieto-occipital IPS and the medial dorsal/ anterior dorsal IPS subregions identified in the structure from motion study (Vanduffel et al., 2002). An alternative explanation for the bilateral superior parietal activations is that they result from spatial- and object-based shifts in visual attention (Corbetta and Shulman, 2002; Kastner and Ungerleider, 2000; Serences et al., 2004; Yantis et al., 2002), which would also be expected during appraisal of these ambiguous stimulus arrays.

4.2.2 Unimodal auditory activations

Auditory areas commonly identified in more than one experiment can be accounted for by similarities in the auditory stimulus features presented to the subjects in each experiment. The primary auditory cortex is located within the medial two thirds of Heschl's gyrus (HG) in humans (Rademacher et al., 1993). Activity within right HG was detected in all experiments and left HG in experiments 1 and 4. These activations reflect the participation of the primary auditory cortex, in auditory object processing (experiment 1), in cyclic pitch sweeps (experiment 2) and in music processing (experiment 4). However, in all experiments the activation of HG appears to extend beyond the medial two thirds into lateral areas, known from previous imaging experiments to be more active during exposure to amplitude modulated (AM) and frequency modulated (FM) sounds than to their unmodulated counterparts (Giraud et al., 2000; Hart et al., 2003). The human superior temporal plane (STP) is commonly subdivided into two main parts: in the planum temporale (PT) posterior to HG and the planum polare (PP) anterior to HG. In addition, cortical areas known to be important in

auditory sensory processing include the ventrolaterally adjacent parts of the superior temporal gyrus and also parts of the inferior frontal gyrus. The differing participation of these auditory areas across experiments reflects differences in the array of acoustic features involved in the stimuli of each experiment. Once again some of the differences in auditory activations across the different experiments may arise due to thresholding effects. Nonetheless, activation differences would be expected to arise due to differences in the specific acoustic features present in the acoustic stimuli of each experiment. Therefore, gross differences in the pattern of unimodal auditory activations across experiments are discussed. Full characterisation of unimodal auditory responses will assist the interpretation of crossmodal integration activations, by enabling us to distinguish between those that are located within and beyond the territories of unimodal auditory cortex.

Unimodal auditory activation of the caudal superior temporal cortex

In experiment 4, in which acoustic stimuli consisted of musical clips, the PT was broadly activated, extending into ventral parietal areas at its posterior extent, whilst in experiments 1 and 2, the PT immediately adjacent to HG was less extensively stimulated. Previous auditory neuroimaging experiments have established that, not only lateral HG, but also PT, were significantly more active in response to FM and AM sounds, than to pure tones or noise (Giraud et al., 2000; Hart et al., 2003). These regions were yet more active when the carrier signal was a harmonic-complex tone rather than a single tone (Hall, 2002). The increased spectrotemporal complexity in experiment 4 and in particular the rapid pitch variations of the various musical stimuli, compared to both the low salience unimodal object sounds (experiment 1) and the simple, pure tone, pitch cycles (experiment 2), may account for the more extensive activation of the PT.

Unimodal auditory activation of the rostral superior temporal cortex

In all three experiments activation of the posterior PP, just anterior to HG, was observed. In experiment 4, this activity progressed rostrally along the STP to the furthest extent, particularly on the right side, where it incorporated anterolateral parts of the superior temporal gyrus (STG) and superior temporal sulcus (STS). In experiment 1, activation of the PP was more extensive on the left than right side and also included anterolateral parts of the STG and STS as it progressed rostrally. In experiment 2, the

excursion into PP and anterior STG was much more extensive on the right, than the left. The increased involvement of the PP and adjacent regions of the right STG have been implicated in previous neuroimaging studies, contrasting melodic versus non-melodic pitch sequences (Patterson et al., 2002; Zatorre et al., 1994), which may explain the extensive activations of these regions in experiment 4. Consistent with this, neurological studies have indicated that right frontotemporal circuits are involved in the perception of global pitch contour (Liegeois-Chauvel et al., 1998; Patel et al., 1998). The observation of right rostral STP activity in experiment 2 suggests that it may have been activated by the rhythmicity of the spectral sweeps inherent in the repeating auditory pitch cycles.

Unimodal auditory activation of the inferior frontal cortex

Extensive activation in the left inferior frontal gyrus (IFG) was observed in experiment 1, while in experiment 2 it was negligible and in experiment 4 there was none. Auditory activations in the left IFG have been noted in previous imaging studies, in which subjects were also required to identify sounds when delivered against a silent background (Lewis et al., 2004), or embedded in one of several natural environmental background sounds (Maeder et al., 2001). The activity at the left IFG can thus be explained as auditory processes involved in the recognition of acoustic environmental sounds, as opposed to the temporal rhythmicity of pitch cycles (experiment 2), or the assessment of auditory aesthetic value (experiment 4).

4.3 The neural correlates of crossmodal integration

The visual-auditory interaction contrast [VA-V>A-Rest] identifies brain areas at which the BOLD response to visual stimulation is significantly positively modulated by concurrent auditory stimulation and *vice versa*. Unfortunately, significant voxels may also theoretically result from this interaction if the BOLD response during the resting baseline condition is greater than during both bimodal and unimodal stimulus conditions. The crossmodal enhancement contrast [VA>V+A] is specifically sensitive to voxels demonstrating increased activity under bimodal versus unimodal conditions, but cannot be used to infer interactions between visual and auditory stimulus processing. A statistical conjunction between the two contrasts, on the other hand, reveals voxels that exhibit both significant response modulation of visual processing during concurrent auditory stimulation (and *vice versa*) *and* increased activity during bimodal versus unimodal conditions. Thus crossmodal integration can be inferred at significant voxels resulting from this conjunction, as both criteria required of the activation profile that would be expected of a brain area that combines visual and auditory sensory information into unified bimodal perception are fulfilled. This has been referred to as the crossmodal integration conjunction, or CMI conjunction.

The CMI conjunction resulted in very different patterns of activation in experiments 1, 2 and 4. In experiment 1, visual and auditory parts of all bimodal stimuli were derived from the same 'object' in the natural environment, thus sharing many intermodally invariant features (described previously - section 1.1.2) and mutually informative with regard to the identity of the 'object' from which they arose. The visual and auditory stimuli of experiment 2, on the other hand, were not intrinsically related in the same manner, as they were abstract, computer-generated stimuli. Temporal covariation could be introduced between cycles of visual random dot array and auditory pitch cycles by matching their rhythmical structure. During synchronous conditions these stimuli were crossmodally correlated, whilst during asynchronous conditions absolutely no mutually informative sensory qualities existed between the visual and auditory stimuli at all. In experiment 4, the pairing of the art stimuli with accompanying musical extracts was entirely arbitrary. As there was no spatiotemporal correlation between the visual and auditory stimuli whatsoever, any visual-auditory interactions / elevations detected were likely to have arisen from simultaneous bimodal processing of unrelated stimuli and would thus reflect generic processes involved in considering

visual and auditory stimuli in parallel. Therefore, the results of the CMI conjunction analysis indicating brain areas at which the visual and auditory stimuli resulting from the same external event were integrated (experiment 1), can be interpreted in light of crossmodal integration activations for physically unrelated stimuli (experiments 2 & 4).

4.3.1 Visual-auditory integration in the primate neocortex revisited

Electrophysiological evidence indicates that superadditive crossmodal responses, so prevalent in the superior colliculus, are very rare when sought in neocortical regions (Benevento et al., 1977). Superadditive responses to concurrent, bimodal stimulation have been detected so infrequently in brain areas outside of the superior colliculus, that its original proponents have revised the originally enthusiastic predictions regarding its likely prevalence in the multisensory neocortex (Stein, 2004). There are several possible explanations for the paucity of support for superadditive indications of cortical CMI. Firstly, very few studies have investigated electrophysiological responses to both unimodal stimuli and concurrent bimodal stimulation. Secondly, CMI cells may be attuned to bimodal stimuli with particular characteristic features, as opposed to synchronously presented pairs of arbitrary 'flashes' and 'beeps' commonly used in electrophysiological studies. Indeed, anecdotal observations in the electrophysiological literature are consistent with this account for the scarcity of CMI activation profiles within neocortical neurons:

> 'a few neurons responded to an object striking a surface, but neither to the sight or sound of the event alone, nor to the simultaneous presentation of a flash and a click' (Bruce et al., 1981)

Two conclusions can be drawn from this statement: a) that responses in some bimodal cells can be elicited *only by combined* visual and auditory stimulation, and that b) some further indication of common cause between the visual and auditory events might be necessary for a response to be elicited. If this is the case, then the rarity of crossmodal interactions at the single cell level might be explained by the inappropriate application of classical electrophysiological approaches to CMI research. The classical approach was to test for neuronal responses with a battery of visual stimuli, followed by a battery of auditory stimuli and then look for crossmodal interactions only in 'bimodal'

cells (defined as responsive to unimodal visual and unimodal auditory stimulation). If the above description of these cells can be taken at face value, then this approach will fail to detect this subset of visual-auditory cells, i.e. those neurons that respond *only* to simultaneous visual-auditory stimulation. Furthermore, if the connectivity of such cells has been attuned through experience to respond to visual-auditory stimulus pairs bound by a common causal event, then probing with arbitrary, discrete, transient visual and auditory stimuli is unlikely to recreate the appropriate conditions.

To expand on this latter point using the example provided by Bruce *et al.*(1981) of an object hitting a solid surface, specific relationships between the spectral profile and the temporal ordering of visual and acoustic events are important in defining the overall multisensory percept. The unimodal visual percept consists of the object moving through space and coming to an abrupt halt. The unimodal auditory percept consists of an acoustic emission with an abrupt onset and offset profile. The crossmodal link between these two unimodal percepts is the cessation of visually perceived motion, which must be coincident with the onset of the acoustically perceived sound, as both result from the same environmental event and are conveyed through different sensory modalities. This bimodal stimulus is defined by more than just the spatiotemporal coincidence between visual and auditory cues, it also requires that the entire sequence of visual motion precedes the acoustic emission. This is because the cessation of visual movement and the onset of the acoustic stimulus are caused by the same event, i.e. contact between the object and the solid surface. Therefore, if 'bimodal' cells exist that are attuned to these kinds of physically constrained visual-auditory relationships, there is a very little chance of recreating such specific stimulus characteristics using brief light flashes paired with simple clicks and tones. It may be that the electrophysiological studies of bimodal stimulus processing may have grossly underestimated the prevalence of CMI sites due to such methodological oversights.

4.3.2 Crossmodal integration for object recognition

Had the desired psychophysical profiles resulted from experiment 1, it may have been possible to demonstrate parametrically varying responses that were correlated with the degree of perceptual enhancement. This would have provided very strong evidence for CMI in the human brain and such a demonstration may well be the only way to definitively prove that CMI occurs at all, as there are several perfectly logical

alternative explanations to account for increased activation during bimodal, as compared to unimodal, stimulus processing. Consequently, in the absence of such a parametric crossmodal recognition enhancement, the inferences formed on the basis of results from the straightforward CMI conjunction analysis are considerably weaker than had been intended. However, that is not to say that this approach is completely without merit, particularly in light of the *a priori* anatomical hypotheses formulated regarding human brain regions implicated by the synaesthesia and neglect conditions which both produce disruption to normal visual and auditory conscious perception (section 2.4).

Past imaging studies investigating visual and auditory object perception

Previous imaging studies of visual and auditory object recognition have generally approached this question from the unimodal perspective. Areas around the border between occipital and temporal lobes at the transverse level of the middle temporal gyrus have been reported to be involved in both visual and auditory object perception. The lateral occipital part of area LOC, which lies posterior and ventral to V5, has proved vital to cue-invariant unimodal visual object processing (Grill-Spector et al., 1998; Kanwisher et al., 1996; Kourtzi and Kanwisher, 2000; Malach et al., 1995; Self and Zeki, 2004), whilst the posterior part of the middle temporal gyrus (pMTG), anterior to V5, has been implicated in auditory object recognition (Lewis et al., 2004; Maeder et al., 2001). The vital question that was addressed in experiment 1 asked where in the brain visual and auditory information interacts during bimodal object recognition. This could have been accomplished either through parallel processing in the visual and auditory object streams, or by integration at putative crossmodal processing centres. Several recent imaging studies have previously attempted to answer these questions (Beauchamp et al., 2004a; Beauchamp et al., 2004b; Laurienti et al., 2003), but have used methodological approaches that seem to be flawed in the following manner.

At least two main strategies exist through which visual and auditory sensory information can be used to identify an object, operating at the sensory level and the semantic level, respectively. As dynamic visual and auditory stimuli arising from the same event unfold over time, mutually reinforcing temporal information is conveyed, helping to reduce sensory uncertainty and increase efficiency by relieving the burden on one or other sensory system, when redundant supramodal information can be extracted. However, if the visual stimulus is static, then there is no possibility of establishing dynamic correlations between visual and auditory information streams, as no intermodal

invariances can be established. In the absence of dynamic correlations in bimodal variation to affirm that the visual and auditory stimuli are in any way related to one another, behavioural enhancements resulting from bimodal, as compared to unimodal, stimulation must therefore operate at a purely semantic level. For instance, it is entirely the prior knowledge derived from experience with telephones that enables stimuli comprising a line drawing of a telephone and the ring of a telephone to result in slightly speeded responses under bimodal, as compared to unimodal, presentation conditions (Laurienti et al., 2003). Only one research group have performed functional imaging studies of *dynamic,* bimodal object processing to date (Beauchamp et al., 2004a; Beauchamp et al., 2004b). They found that an ROI-defined area preferring complete visual and auditory objects to their scrambled counterparts, located in various different regions of the posterior middle temporal gyrus / posterior superior temporal gyrus (pMTG/pSTG) in different subjects, was also more active during bimodal than unimodal object identification. However, this study used a contrast that revealed areas where greater BOLD activity was detected during bimodal, as compared to unimodal, activations. In the absence of a resting baseline condition, they could not perform a visual-auditory interaction contrast and so could not infer that visual processing was modulated by concurrent auditory processing, and *vice versa.*

The posterior temporoparietal junction in crossmodal integration

The CMI conjunction of experiment 1 revealed activation of the posterior temporo-parietal junction (pTPJ). The pTPJ lies posterior and dorsal to the site(s) implicated by the Beauchamp study. Whether or not the activity detected at the pTPJ truly reflects CMI for object recognition is a matter for debate, as it is very close to the temporo-parietal junction (TPJ) activations implicated previously in studies of bottom-up attention (Corbetta and Shulman, 2002; Downar et al., 2000; Downar et al., 2001; Downar et al., 2002). The term pTPJ is used rather than TPJ, to emphasise the fact that it is displaced in a posterior direction, as compared to the coordinates given for TPJ activations in these studies of multisensory stimulus salience.

Imaging studies of visual attention have isolated the sources of attentional control signals to specific regions of posterior parietal and frontal cortex, along the intraparietal and superior frontal sulci, respectively. These observations were based on experimental paradigms designed to investigate spaced-based, feature-based and object-based attention (Beauchamp et al., 2001; Corbetta et al., 2000; Giesbrecht et al., 2003;

Hopfinger et al., 2000; Le et al., 1998; Liu et al., 2003; O'Craven et al., 1999; Serences et al., 2004; Shulman et al., 2002; Vandenberghe et al., 2001; Wojciulik and Kanwisher, 1999). In each of these studies of visual attention, subjects were cued to redirect their attention to a certain part of space, a certain feature, or a certain object. Such endogenous, subject-driven, voluntary attentional shifts appear to be controlled by dorsal parietal and frontal areas, without invoking activation in the TPJ. However, when attentional shifts are stimulus-driven, or in the popular parlance of the attentional literature - 'bottom-up' - activations in the vicinity of the TPJ are invariably elicited, regardless of whether the salient stimulus transition occurs in the visual, auditory or tactile sensory modality (Downar et al., 2000; Downar et al., 2001; Downar et al., 2002). Therefore, the CMI response profiles that have been isolated at the anterior TPJ (e.g. experiment 2) may simply be explained by an increased frequency of salient features detected during bimodal conditions (in response to both visual *and* auditory sensory stimuli), compared to unimodal conditions (where only visual or auditory sensory events occur, thus inducing relatively fewer transient activations).

The right-lateralised cluster of pTPJ voxels at which CMI response profiles were also identified in experiment 1 [48, -62, 32, MNI], lie posterior and dorsal to those found in the Downar studies (Downar et al., 2000: [54, -42, 13, MNI]; Downar et al., 2001: [58, -44, 16, MNI]. A translation of 20mm in both directions might be deemed negligible and in the absence of an explicit control to enable the dissociation of responses to attentional cues from recognition cues, would not stand up to criticism. However, as V5 and LO are separated by an even smaller distance (Beauchamp, 2005; Downar et al., 2001; Self and Zeki, 2004), yet are specialised to extract entirely different visual features, this difference is nevertheless worth pointing out. Furthermore, the notion that sites close to the TPJ may be involved in more than just attentional modulation has been raised previously, as they have been implicated in imaging studies investigating a broad range of social cognition tasks (Allison et al., 2000), moral judgement (Greene and Haidt, 2002) and theory of mind (Gallagher et al., 2000; Greene and Haidt, 2002). In addition this activation lies just posterior to the TPJ lesion sites implicated in neglect (Vallar, 2001) and lateral to the parietal-occipital junction activation of the Paulesu synaesthesia study (Paulesu et al., 1995).

In light of these facts, it seems reasonable to speculate about the underlying causes of visual-auditory interactions and bimodal elevations detected in the right pTPJ in this experiment. Significant voxels from the CMI conjunction within the right pTPJ

may be the site at which CMI integrates visual and auditory sensory information, possibly by coregistering visual and auditory stimuli in space (Andersen, 1997). The synchrony between visual motion events embedded in visual noise and the auditory events embedded in acoustic white noise, may enable the common cause (i.e. the 'object' in question) from which both stimuli arise to be established. This, in turn, may have the effect of decreasing the overall amount of information to be processed (Lewkowicz, 2000), increasing the salience of the crossmodally-integrated percept and thus improving recognition under bimodal conditions. However in the absence of convincing performance data to demonstrate such recognition improvements under bimodal conditions, this interpretation must remain speculation. All that can be concluded with any confidence, is that voxels within the right posterior TPJ, neither implicated in unimodal contrasts nor previous 'bottom-up' attentional studies, demonstrate an activation profile consistent with crossmodal integration of visual and auditory stimuli in an object recognition task.

4.3.3 Crossmodal integration between simple, cycling stimuli

The CMI conjunction performed in experiment 2 must be viewed in a different light to the other experiments, as crossmodal synchrony or asynchrony only exists under bimodal conditions, i.e. it is absent during unimodal conditions. In the other two experiments, object recognition (experiment 1) and aesthetic rating (experiment 4) was possible under bimodal and unimodal conditions, which is a more appropriate context in which to evaluate visual-auditory interactions. Evaluating the CMI conjunction for synchronous stimuli and asynchronous stimuli resulted in near identical activation profiles. This was surprising in light of the Calvert group's experiment (Calvert et al., 2001), which emphasised the reliability of response enhancements resulting from synchronous bimodal versus unimodal stimulation and of response depression induced by asynchronous bimodal versus unimodal stimulation. Collapsing across synchronous / asynchronous bimodal stimuli, several areas of significant visual-auditory interaction and bimodal elevation were identified, Of these, only the right-lateralised anterior temporoparietal junction cluster [52,-50,22, MNI], demonstrated parameter estimate differences indicative of a positive visual-auditory interaction *and* an elevated response to bimodal stimulation in comparison to the unimodal responses. The coordinates of this activation correspond to the anterior TPJ activation identified in a series of imaging

experiments, which responded to multisensory stimulus salience induced by changes in continuously presented visual, auditory and tactile stimuli (Downar et al., 2000; Downar et al., 2001; Downar et al., 2002). As the activation in experiment 2 appeared to overlap with that of the Downar experiments, it is likely that the CMI activation profile could be explained by the detection of an increased incidence of stimulus changes when visual and auditory stimuli are presented together, as opposed to apart.

4.3.4 Crossmodal integration for aesthetic evaluation

Crossmodal integration conjunctions were investigated for stimuli that were on average considered slightly more beautiful during bimodal than unimodal conditions and for those stimuli in which the bimodal ratings exhibited no such enhancement. The aim was to isolate the neural correlates of positive crossmodal beauty interactions. Unfortunately, the results were not conducive to such an interpretation as both contrasts revealed essentially the same distribution of significant voxels. This may suggest that the aesthetic processing of arbitrarily related visual and auditory stimuli is not significantly modulated under bimodal conditions. An alternative explanation is that the crossmodal enhancements observed in this study were either not large enough, or not consistent enough across subjects, to produce a detectable interaction and response enhancement.

Collapsing across the aesthetically enhanced and non-enhanced bimodal pairs, several cortical areas were revealed at which significant visual-auditory interactions and bimodal elevations could be observed. Two sites of CMI were identified along the right STS. The first of these activations consisted of 384 voxels with the maximally significant voxel (Z=5.25) midway along the length of the STS [54,-10,4, MNI] including the voxel identified in the Paulesu synaesthesia study [64,-14,-10, MNI] (Paulesu et al., 1995), detailed in section 2.4.1. The other STS activation was located posteriorly in the immediate vicinity of the pSTS/MTG, at which bimodal elevations have been identified in previous studies of visual-auditory integration (Beauchamp et al., 2004b). That CMI activation profiles have been found in experiment 4 at a very similar site to that of the Beauchamp study, casts doubt on the assertion that this region is involved in integrating object-related bimodal information for related, but not unrelated pairs. This is because there is neither physical nor even semantic correspondence between the visual and auditory stimuli in experiment 4. On the other

hand, recent evaluation of multisensory responses at this locus using the ultra-high resolution of parallel fMRI indicates that within this area, a 'patchy' organisation of visually responsive, acoustically responsive and bimodally responsive cortex exists (Beauchamp et al., 2004a). Therefore, it is possible that the CMI conjunction registers significant increases in both the visual and auditory patches during concurrent visual and auditory stimulation, despite the absence of congruence between them that would invoke the additional participation of intervening multisensory patches, as many of these subdomains occupy the volume of a single voxel.

Activity in a part of the anterior cingulate (AC) and ventral occipital areas was also identified using the CMI conjunction in the aesthetics study. These areas have been previously observed to induce stronger BOLD responses under semantically congruent bimodal pairings of a visual line drawing (e.g. a cow) with the matching acoustic stimulus (e.g. a 'moo' sound), than when the image was paired with a non-matching acoustic stimulus (e.g. a 'quack' sound) (Laurienti et al., 2003). Significant visual-auditory interactions and bimodal enhancements were identified in experiment 4 at both these areas, despite the absence of demonstrable stimulus congruence, suggesting that visual and auditory information may be integrated at this site irrespective of any specific semantic relationships. In the Laurienti study neither resting nor unimodal stimulus conditions were included, precluding comment on any CMI-related activity. The observations of the Laurienti study, stressing the sensitivity of these areas to crossmodal congruence are not compatible with the observations of visual-auditory interactions and enhancements regardless of the lack of demonstrable crossmodal congruence in experiment 4. If the cause of these elevated responses is unrelated to semantic congruence between stimuli, then perhaps it is simply a question of stimulus salience. While the congruent stimuli would have been more salient than the incongruent pairs in the Laurienti study, in experiment 4 the bimodal stimuli would have been more salient than the unimodal stimuli.

In addition to these areas previously identified in neuroimaging studies of crossmodal integration (Beauchamp et al., 2004a) and bimodal contrasts (Laurienti et al., 2003), several areas were also identified that have not previously been identified in previous neuroimaging studies investigating visual-auditory integration. These include activations in the posterior cingulate and the left supramarginal gyrus, which might be accounted for by increased demands on attentional resources (Corbetta and Shulman, 2002), but also the mid-right STS, an area not generally thought to be involved in

attention. Activation of the mid-STS has been implicated in the perception of biological motion in human imaging studies (Allison et al., 2000). Recent primate electrophysiological investigations of multisensory integration have implicated the primate mid-STS in visual and auditory integration of observed biological motion (Barraclough et al., 2005). Yet there is no biological motion in the art and musical stimuli of experiment 4. In a neuroimaging study of visual-auditory speech processing, Wright and colleagues (Wright et al., 2003) identified increased BOLD activity at the right mid-STS, but again in experiment 4 there were no speech stimuli. These observations suggest that this area may in fact be another generic site of CMI, more active during bimodal than unimodal visual/auditory stimulation, irrespective of any physical and/or semantic congruence inherent in the stimulus pair.

4.4 Crossmodal and intramodal synchronicity

4.4.1 Improving on previous studies of crossmodal synchronicity

Previous attempts to use functional imaging studies to investigate the neural correlates of visual-auditory synchrony and asynchrony using non-speech stimuli have resulted in incompatible findings (Bushara et al., 2001; Calvert et al., 2001). The root cause of this incompatibility lies in the assumptions that were made regarding the expected response profiles and implicit in the design of their experimental paradigms.

The fMRI study of Calvert and colleagues (Calvert et al., 2001) based their entire methodological approach on the electrophysiological rules of visual-auditory integration in the cat superior colliculus (Meredith and Stein, 1986). According to these principles, it was assumed that synchrony between brief, discrete, acoustic white noise bursts and reversals of a rapidly alternating (8Hz) checkerboard stimulus, would induce response enhancements, whilst asynchrony would elicit response depression. These assumptions effectively rendered the study totally blind to the possibility of increased responses during asynchronous versus synchronous conditions.

Meanwhile at the National Institute of Health, other researchers were specifically investigating the neural correlates of crossmodal asynchrony, by exposing subjects to brief, momentary flashes of light and acoustic bleeps (Bushara et al., 2001). The approach was to parametrically manipulate the temporal offsets between visual and auditory events in order to cover a range of stimulus onset asynchronies for light leading sound and sound leading light. Brain areas were then sought at which the magnitude of activity was positively correlated with the difficulty of asynchrony perception, i.e. greater activity during slighter intervals between visual and auditory stimulus onset. They also performed a subtraction analysis contrasting asynchronous against synchronous stimulus conditions, but did not report the reverse contrast, so any voxels more active during crossmodal synchrony than asynchrony were not revealed. In experiments 2 and 3 the neural correlates of crossmodal synchronicity were approached in the most impartial way possible, so that the eventual findings would not be biased by specific expectations. This was realised through the use of the factorial design outlined in section 3.4.5.

From a theoretical perspective there is a further issue, vital to the proper ordering of multisensory perception, which has not previously been addressed in the context of visual-auditory perceptual integration. The co-occurrence of unrelated visual and auditory events can occur quite frequently in busy environments, so synchrony detection alone is not necessarily sufficient to support robust crossmodal integration. Thus, a crucial mechanism required by a system that integrates information across the sensory domains, is to ensure that chance co-occurrences of visual and auditory stimuli are not mistaken for causally linked emissions from the same event. However, a consistent correlation between visual and auditory events, even over a relatively short period of time, can be strongly indicative of common cause and can thus be used to identify suitable cues for perceptual unification. Once a consistent correlation between visual and auditory variation over time has been established, it is then appropriate for the sensory signals to be integrated, but only once the potentially dangerous risk of integrating totally unrelated sensory pairs has been adequately ruled out. The motivation for choosing continuously varying stimuli in experiment 2 was to permit the perceptual judgement of synchrony to rely upon detection of consistent visual-auditory co-variation, during a relatively prolonged exposure to the pair of dynamic, bimodal, sensory 'objects', rather than the simple onset matching between discrete, static visual and auditory events used in other studies.

4.4.2 Brain areas exhibiting significant responses to synchrony

In trying to predict the likely regional BOLD responses to synchronous and asynchronous stimulus conditions, logic and the literature suggested two feasible outcomes. Responses to visual-auditory synchrony as opposed to asynchrony was expected either to result in increased activity: 1) in both visual and auditory unimodal cortices, which would be indicative of increased levels of inter-sensory communication, or 2) in separate regions of 'association cortex' beyond the unimodal territories.

The absence of synchrony related activity in experiment 2

In experiment 2, no regions of increased activation under conditions of bimodal synchrony (SYNC) versus asynchrony (ASYN) were observed anywhere in the brain at corrected significance. This is surprising given that the Calvert synchrony study found several areas more active under synchronous than asynchronous conditions, including

the insula bilaterally, superior colliculus, thalamus, superior temporal sulcus, intraparietal sulcus and ventromedial and dorsal prefrontal cortex (Calvert et al., 2001). There are several differences between the stimuli used in the Calvert study and experiment 2 that could potentially account for the differences.

Firstly, psychophysical studies have suggested that when visual stimuli lead auditory stimuli by up to 0.15s, subjects consistently perceived them as simultaneous (Stone et al., 2001). In the Calvert study even if the visual and auditory stimuli were presented in counterphase, the stimulus onset asynchrony would be only 0.0625s, less than half the time required for asynchrony to be consciously perceived according to the findings of Stone and colleagues. It is therefore quite possible that subjects could not, in fact, consciously discriminate between the synchronous and 'asynchronous' stimulus conditions in the Calvert study. If this were the case, then the activations that they found might at best be interpreted as the neural correlates of 'subconscious' synchrony detection and so would not be implicated in experiment 2 as the synchrony and asynchrony could be easily distinguished (96.3% mean correct responses).

Secondly, stimuli in the Calvert study were discrete momentary events and so the synchronous condition may induce stronger BOLD activity in areas sensitive to matching versus non-matching, asynchronous visual and auditory onset times. In experiment 2 however, the onsets of visual and auditory stimuli were always simultaneous and the synchrony or asynchrony between these continuously cycling stimuli was resolved according to the presence or absence of a consistent temporal covariation between these bimodal sensory streams. The brain areas implicated in the Calvert study may therefore reflect the neural correlates of discrete, bimodal onset matching, whilst the absence of onset mismatches in the asynchronous condition in experiment 2 might explain the absence of significantly increased responses in the synchrony contrast [SYNC > ASYN].

Thirdly, whilst the visual and auditory stimuli in the Calvert study were presented rapidly every 0.125s, the stimuli in experiment 2 cycled slowly over a period of between 1.2-3.0s. Therefore an alternative, or additional, explanation is that the high rate of stimulus delivery induced stronger activity in the synchrony-sensitive brain areas in the Calvert study. The stimulus cycles in experiment 2 occurred only once for every 9-24 bimodal events in the Calvert study and so may have been too infrequent to induce significantly different responses at these same neuroanatomical sites.

Brain areas demonstrating significant responses to synchrony in experiment 3

In experiment 3, the crossmodal synchrony contrast [SYNC>ASYN] revealed significant activation only at the amygdala bilaterally. The analogous intramodal synchrony contrast revealed significant activation elsewhere at the left angular gyrus and the parietal-occipital sulcus between cuneus and precuneus, but not in the amygdala. This suggests that the amygdala is specifically responsive to crossmodal temporal covariation. However as no significant activations were identified in the crossmodal synchrony contrast in experiment 2 the several stimulus differences between the two experiments will be outlined.

In experiment 3 there were two pairs of bimodal stimuli, whilst in experiment 2 there was a single pair of bimodal stimuli. Therefore, the activation of the amygdala in the latter experiment may be involved in specifying which of the two possible visual-auditory pairs were synchronised with each other. A second difference in stimulus features was the continuously varying rhythmic stimuli in experiment 2 in contrast to the discretely varying, randomly occurring stimuli of experiment 3. Therefore an alternative explanation is that the amygdalar activations may be specifically responsive to synchrony between unpredictable timing of visual stimulus movement from quadrant to quadrant when coincident with the discretely cycling acoustic events. Finally the critical difference may be related to the rate of stimulus delivery, as stimulus events occurred more rapidly in experiment 3 (every 0.32-1.36s) than experiment 2 (1.2-3.0s).

Equally the observation of increased amygdala activation during crossmodal synchrony in experiment 3, but not in experiment 2, could be attributed to the different numbers of stimuli used in each experiment. In experiment 2 there was only one possible crossmodally synchronous configuration (between the one visual and one auditory stimulus), whereas in experiment 3 there was two possible configurations (between the two visual and the two acoustic stimuli). The role played by the amygdala in experiment 3, but not in experiment 2, may thus involve perceptual unification of the synchronised pairs of visual and auditory cues when more than one combination was possible. Once again no evidence was found to support the observations made in the Calvert study, and neither did they report differential amygdalar responses.

The structure and function of the amygdala

Anatomically the amygdala is an extremely complex cluster of numerous nuclei that often merge with their neighbours, as well as with adjacent non-amygdalar regions.

These are generally divided into three main groups: the basolateral nuclear group, the superficial cortex-like nuclear group and centromedial nuclear group (McDonald, 1998). Electrical stimulation of the different subdivisions elicits different behaviours according to the structures targeted by the efferent projections, e.g. the basolateral group project to the striatum and are thought to be critical for voluntary, instrumental behaviour, whilst the central group project to the brainstem and hypothalamus and may be critical for reflexive, classically conditioned responses (Killcross et al., 1997). Stimulation and ablation studies in experimental animals have shown that the amygdala is in fact involved in a wide range of behaviours related to biological drives and motivation including: arousal, orienting, sleep, fight-or-flight, feeding, drinking, reproductive, maternal, reward and punishment (Kaada, 1972). Yet functional imaging studies investigating amygdala function invariably concentrate on its role in the acquisition of associations between emotionally salient stimuli such as fearful faces and loud aversive noises (e.g. Buchel et al., 1998; Morris et al., 1998). The incentive for this is understandable, as ever since Kluever-Bucy syndrome was described in monkeys, by the scientists after whom this condition was named (Kluever and Bucy, 1939) and others before them (Brown and Schafer, 1888), various adverse effects on normal emotional behaviour were immediately obvious. Human lesion studies have since demonstrated, in agreement with the extensive animal literature that further scrutinised amygdalar function since those early studies (reviewed in: LeDoux, 1995), that aversive conditioning is negatively affected by amygdalar damage (Bechara et al., 1996; LaBar et al., 1995). Despite the undeniable involvement of the amygdala in fear conditioning, as described above, it is just one of many important roles played by this complex structure. Stimulation of the human amygdala in patients with chronic temporal lobe epilepsy results in a diverse array of experiences, including both positive and negative emotional responses, but also visual and auditory hallucinations (Gloor et al., 1982). It seems that the role of the amygdala in perception of non-emotional, but nonetheless salient, biologically relevant, stimuli has been somewhat overlooked in favour of its compelling role in the acquisition and expression of conditioned fear responses (LeDoux, 1996).

This is the first time, to our knowledge, that increased functional activation of the amygdala has been implicated by simple manipulation of temporal covariation between bimodal sensory stimuli. Previously, an inability to perform visual-tactile crossmodal associations has been documented in primate lesion studies, as a result of

bilateral amygdalectomy (Murray and Mishkin, 1985). However, subsequent excitotoxic lesion studies indicated the involvement of adjacent perirhinal cortex, rather than amygdala in sensory-sensory associative learning, suggesting that previously reported deficits on cross-modal tasks were probably due to inadvertent damage to perirhinal cortex during surgical removal of the amygdala (Goulet and Murray, 2001). Nevertheless other research has suggested that an intact amygdala is crucial for intermodal associative learning involving the biological value of stimuli, but not for maintaining the value of secondary reinforcers once they have been learned (Malkova et al., 1997). It is difficult to use this account of the amygdala's role in the formation of crossmodal associations to explain its involvement in experiment 3 in which the temporal relationships between the visual and auditory stimuli changed from trial to trial. In the absence of a convincing explanation for how crossmodal associations might be useful in an experimental paradigm where any such associations formed on the basis of one trial would immediately be undone in the following trials, the most logical alternative is that the amygdala is simply sensitive to crossmodal synchrony.

Should the list of functions attributed to the amygdala be extended?

Anatomical projections from rostral parts of the visual and auditory cortical domains of the temporal pole to the amygdala have been extremely well characterised (Romanski and LeDoux, 1992; Turner et al., 1980) and of the numerous amygdalofugal projections, direct influence over the primary visual and auditory cortices are present in the rat, cat and primate (McDonald, 1998). Furthermore, the main output of the superior colliculus is to the posterior thalamic pulvinar nucleus (Benevento and Fallon, 1975), which also has direct reciprocal connections with the amygdala (Amaral, 1992) and has been proposed as the substrate through which fearful faces are processed unconsciously (Morris et al., 1999). However a closer examination of Morris's study reveals a conspicuous confound, whereby the masked (unconsciously perceived) fearful face stimulus is paired with an auditory tone (i.e. synchronous bimodal stimulation), whilst the unmasked (consciously perceived) stimulus is not (unimodal stimulation). Therefore, in light of our findings, the tecto-thalamo-amygdalar pathway that they propose as the neural substrate for unconscious processing of emotive faces, may in fact simply be conveying information regarding the synchronicity between concurrently presented visual and auditory stimuli, irrespective of the emotional content.

In a recent meta-analysis of 55 PET and fMRI investigations of emotional stimulus processing (Phan et al., 2002), amygdala activations were found to occur in response to various evocative stimuli, including fearful faces, sad faces, happy faces, aversive pictures, positive pictures and even nonaversive/neutral pictures, suggesting that the amygdala may not exclusively respond to affectively laden stimuli, but to meaningful stimuli in general. The amygdala may play a more general role in enhancing information processing to any salient stimuli, whether aversive or nonaversive (Everitt et al., 1991) and regardless of emotional valence (Lang et al., 1993). If this is the case, then perhaps the amygdala operates along a continuum of salience, with emotional and particularly aversive stimuli at one end of this continuum as the most salient of biologically relevant stimuli, whilst the elevated perceptual salience arising from dynamic covariation between non-emotive bimodal cues are also processed at this structure.

This is the first time that functional activation of the bilateral amygdala has been recorded in humans during a non-emotive visual-auditory crossmodal perceptual task, which perhaps calls for the above re-evaluation of its role in perception. The classical literature on amygdala function would not have predicted an elevated response to crossmodal synchrony between pairs of non-emotive stimuli, although the potential for this structure to provide, "fascinating insights into yet unknown forms of multisensory convergent circuitry," has been noted previously by distinguished scientists in the field of crossmodal integration (Meredith, 2002). The most likely explanation should incorporate the dominant role of the amygdala in crossmodal conditioning, where it is instrumental in forging predictive links between pairs of sensory cues according to their simultaneous or consecutive detection. As tight crossmodal covariation of the like exhibited in experiment 3 strongly indicates a common cause driving both visual and auditory stimulus transitions, perhaps this translates into increased salience, inducing a consequential elevation in amygdala activity, which may forge a (transient) predictive link between the two.

4.4.3 Brain areas exhibiting significant responses to asynchrony

In experiments 2 and 3, the same area of right anterior frontal operculum (aFO) was found to be more active under conditions of crossmodal asynchrony than synchrony, despite considerable differences in the experimental paradigms used. Experiment 3 both confirmed the main finding of experiment 2 and also enabled the inference that it is specific to crossmodal asynchrony, as aFO activity was not implicated in the intramodal asynchrony contrast. The main effect of intramodal asynchrony revealed significant activity in both superior temporal gyri and V5, bilaterally. This suggests that the visual-visual intramodal asynchronies between wedge movements probably induced the V5 activity, given that it is generally accepted to be involved in visual motion processing, whilst the auditory-auditory intramodal asynchrony between the two instrumental note sequences induced the activity in superior temporal gyri, as it borders known areas of unimodal auditory cortex. The repeated observation of right aFO activity specifically under crossmodally as opposed to intramodally asynchronous conditions is arguably the most important finding of this thesis. It indirectly suggests a putative role for this region in the regulation of visual auditory integration for unified perception.

Crossmodal asynchrony and the right anterior frontal operculum

In experiment 2, a continuous cluster of voxels was identified connecting the right anterior short insular gyrus and the anterior frontal operculum (aFO, neuroanatomical location outlined in detail in section 4.6) that were significantly more active ($p<0.05$, corrected) during ASYN than SYNC bimodal stimulus epochs. This was in general accordance with the findings of Bushara and colleagues (Bushara et al., 2001), who found that activity varied parametrically with the degree of bimodal asynchrony throughout the right anterior insula [37,23,-5, Talairach], directly adjacent to the deepest activation in experiment 2 [36,18,2, MNI], which approximates very closely to the equivalent Talairach coordinates [38,18,1, Talairach]. As outlined above, the design of the Calvert study prohibited the discovery of voxels that were more active under asynchronous than synchronous conditions (Calvert et al., 2001). Interestingly, significant activity within the anterior insula, at very similar coordinates to the anterior insular activation in experiment 2, was also identified in a recent fMRI study investigating interoception, but was given a different interpretation (Critchley et al.,

2004). This involved the subjects judging whether their own heartbeat was in or out of synchrony with an acoustically relayed tone sequence, suggesting that the right anterior insula may be sensitive to asynchrony between bimodal stimuli across a variety of different sensory modalities.

The involvement of the aFO in asynchrony detection in both experiments 2 and 3 is an entirely novel finding. Although in the subtraction analysis and subsequent correlation analysis of the Bushara study significant voxels were identified in the right prefrontal cortex, these were displaced rostrally with respect to the insula, away from the aFO. The involvement of the right aFO may rely upon the dynamic nature of the bimodal stimuli used in our experiments and implicated only when visual and auditory stimulus cycles progress continuously (experiment 2) or discretely through a sequence of stimulus states (experiment 3), but not when simple, static visual and auditory stimuli are used (Bushara, 2001). This may amount to the difference between identifying crossmodal asynchrony based on departure from cyclical covariation, rather than departure from simple stimulus onset matching; a temporally global rather than local processing mechanism. This topic will be revisited in the last section (section 5.6).

Factorial interactions between synchronicity and other factors

No significant interactions were found in experiment 2 between relative synchrony (AvS) and the visual submodality through which the visual changes were conveyed (colour versus motion) anywhere in the brain. Thus there is reason to believe that the right aFO neurons are sensitive to asynchrony between visual and auditory stimuli regardless of visual submodality. This notion is also supported by the involvement of the right aFO in experiment 3, where visual stimulus changes were in quadrant position and discrete cycles through four different musical notes. The invariance to visual sub-modality is important, as it implies a generic role in crossmodal asynchrony detection regardless of which particular visual attributes change. This would be expected of any brain area involved in the regulation of multisensory grouping, in order for it to perform its function for all possible bimodal stimuli encountered in the natural environment.

In experiment 2, no significant voxels were detected in the interaction of stimulus synchronicity and regularity (irregular versus regular) within voxels identified in the asynchrony contrast. This indicates that the predictability of the signal did not affect asynchrony related activity at the right aFO. However, a significant interaction

between synchronicity and regularity *was* detected in a small cluster of voxels adjacent to the anterior insula subcluster. The parameter estimates at the maximally significant voxel within this cluster, indicate that the difference in the magnitude of the response under asynchronous versus synchronous conditions was significantly larger for stimuli of regular, as compared to irregular, cycle sequences. This may indicate that asynchrony is more readily detectable at this locus when the visual and auditory stimuli are predictable in their temporal cycle pattern, than when unpredictable. This did not appear to affect the voxels identified in the crossmodal asynchrony contrast as these were located laterally in the overlying aFO.

4.5 Visual-auditory neuroaesthetics

4.5.1 Studying beauty

From the Platonic perspective, beauty is the recognition, or abstraction, of the "eternal forms of goodness, harmony and symmetry". Such descriptions lend themselves equally well to the exaltation of both visual and auditory sensory artistic phenomena. Until recently, the scholarly analysis of the uniquely human faculty of aesthetic appreciation has largely been the preserve of artists, art historians and great philosophers. Not until the lucrative advent of modern cosmetic surgery did the first surge of serious scientific analysis of what defines a beautiful face (Broadbent, 1989; Romm, 1989), smile (Snow, 1999), and body (Leist, 2003), get underway. Implicit in these themes is our preoccupation with human physical beauty. It is therefore not surprising that the inaugural study of neural correlates of beauty sought brain regions underpinning the perception of beauty in human faces (Nakamura et al., 1998). This theme has more recently been extended to encompass the enhanced appeal of a smiling, attractive face (O'Doherty et al., 2003). Activation of the medial orbitofrontal cortex (OFC) accompanied perception of an attractive as opposed to an unattractive face, which was further augmented by a smiling facial expression. In such studies the attractive visual object in question was always a human face and so activation of the medial OFC may be related to other factors that covary with attractiveness, such as reproductive fitness (Thornhill and Gangestad, 1999) and social relevance (Allison et al., 2000). Furthermore, powerful, instinctual drives and emotional influences may dominate the perception of attractiveness in a face and may therefore be regarded as different from an analytical, cognitive judgement of beauty.

The aim in experiment 4 was to examine the neural correlates of an aesthetic evaluation achieved through unhurried, elemental analysis of the complex inter-relationships between stimulus features such as symmetry, harmony and juxtaposition. It was hypothesised that a considered, aesthetic beauty judgement of this nature might well be expected to recruit brain areas distinct from those responsible for assigning motivational reward value to highly emotive auditory stimuli (Blood and Zatorre, 2001) or consumables such as food and narcotic drugs (Berridge, 1996; Breiter and Rosen, 1999; Rolls and Baylis, 1994; Small et al., 2001; Stein et al., 1998; Tremblay and Schultz, 1999).

4.5.2 Recent advances in neuroaesthetics

Recent human neuroimaging studies have replicated findings from the animal literature, emphasising the role of medial OFC, amygdala, ventral striatum and medial prefrontal cortex in reward (for recent review: McClure et al., 2004). In addition, visually perceived images of differing emotional valence result in differential activation profiles in ventrolateral, dorsomedial and dorsolateral prefrontal cortex (e.g. Dolcos et al., 2004). The judgement of aesthetic beauty could easily be obscured beneath the inter-related, yet separable, neural correlates of emotion and reward. In experiment 4 a specific attempt was made to de-emphasise these potentially confounding influences by explicitly instructing subjects to make a 'considered judgement' as to how beautiful they found a selection of visual and auditory artworks.

For centuries, artists and musicians alike have spent their lifetime pursuing the ultimate objective of creating works of great beauty. As such, their ability to induce 'beautiful sensations' in the minds of their audiences effectively makes them a type of sensory neurobiologist, as well as an artist (Zeki, 1999). Thus, the most effective tools with which to probe the neural correlates of aesthetic beauty must surely be the great work of accomplished artists and musicians.

In the fMRI study of Kawabata and Zeki (Kawabata and Zeki, 2004), subjects viewed portrait, landscape, still life or abstract compositions that they considered to be beautiful, compared to those that they considered neutral or ugly. Direct comparison of these conditions yielded significant voxels in the medial OFC. In an earlier positron emission tomography study, Blood & Zatorre required subjects to listen to sounds that varied along a continuum of consonance to dissonance in order to identify regions that correlated with the directly related perceived 'pleasantness' of the sound (Blood et al., 1999). They also found increased activation in the medial OFC to consonant versus dissonant sounds. The common involvement of medial OFC in both visual and auditory studies could indicate that the attractiveness of artistic images and sounds might be processed at this site in a supramodal fashion, that is, independently of the sensory modality through which the information was conveyed. However, the specific location of these medial OFC activations differed across studies.

The visual art study contrasted exposure to subjectively beautiful versus ugly stimuli, yielding significant activation in left mOFC, whilst the auditory study revealed

a primarily right-lateralised and caudal region of mOFC in which activity correlated with sound 'pleasantness'. This may reflect either modality specific processing differences, or differences in the specific instructions given to subjects regarding how they should processes and respond to the stimuli. However, as no previous study has ever investigated the aesthetic beauty judgement for visual and auditory stimuli in the same subjects in a single experiment, no consensus has yet been reached as to which brain regions are involved in the modality-invariant aesthetic judgement of beauty.

The pursuit of brain areas differentially responsive to the degree of beauty in both visual artworks and acoustic musical extracts is not only interesting with regard to neuroaesthetics. In the context of the aims of this thesis it is also highly relevant with regard to how visual and auditory information is extracted beyond those brain areas that are specifically involved in unimodal stimulus processing. Demonstration of modality-invariant regional brain responses to a specific attribute shared by visual and acoustic stimuli such as beauty, yet with an implicit reliance upon unimodal processing as a prerequisite for the aesthetic evaluation, also serves to demonstrate brain areas at which unimodal stimuli are evaluated at the supramodal cognitive level.

4.5.3 Role of the lateral orbitofrontal cortex in aesthetic evaluation

During both unimodal visual art and unimodal musical aesthetic evaluation, significant correlations with the 0^{th} order term, i.e. irrespective of the actual aesthetic rating given, were observed at the left intra-parietal sulcus (IPS) and the left lateral orbitofrontal cortex (latOFC). This was formally confirmed through a conjunction analysis ($p<0.05$, corrected). Whilst the activity in the left IPS may well be explained by neural mechanisms supporting sustained attention to the sensory stimuli during beauty evaluation, as compared to the implicit baseline (Corbetta et al., 1991; Corbetta and Shulman, 2002; Giesbrecht et al., 2003; Kastner and Ungerleider, 2000; Liu et al., 2003; Serences et al., 2004; Vandenberghe et al., 2001), a different explanation is required for the latOFC activation. The involvement of the latOFC, a brain region not usually implicated in attentional processing, is better explained by a role in the refinement of the aesthetic value assignment, i.e. establishing the most suitable score for each particular piece. Such a role for the lateral OFC has been identified in previous imaging experiments investigating reward, motivation and goal selection and supports this interpretation (review: Elliott et al., 2000a).

In a recent PET experiment (Arana et al., 2003) designed to separate the neural substrates of incentive values from goal selection, subjects were required to make a selection from a menu in some trials, or merely view the menu in others. Medial OFC activation was detected when confronted with high incentive menus and when making a choice between items, whilst right lateral OFC was only observed when subjects had to make difficult selections between many enticing options. This can be interpreted as the instinctive evaluation of the stimulus by the medial OFC, whilst the more considered, careful evaluation of the several appealing food choices invokes the participation of the lateral OFC. Several other experiments support this dissociation between selection of appropriate 'choice' at the lateral OFC and automatic stimulus valuation at the medial OFC (Elliott et al., 2000a; Elliott et al., 1999a; Elliott et al., 2000b; O'Doherty et al., 2001; Small et al., 2001). However in these studies the right latOFC is invariably implicated, but rarely the left. Therefore, the left OFC may be specifically activated during the 'choice' of the most fitting score that suits the aesthetic appeal of a visual artwork or musical extract, rather than the arbitrary selection from a list of items (Arana et al., 2003) or the selection of a favoured 'guess' (Elliott et al., 2000a).

4.5.4 Instinctive versus considered beauty evaluation

The most important observation in experiment 4 was revealed by the 1st order parametric regression analysis. A cluster of voxels at the medial aspect of the inferior transverse frontopolar gyrus was identified at which the magnitude of the BOLD response was positively correlated with each subject's own perceived beauty ratings, regardless of whether visual or auditory stimuli were being evaluated (figure 41).

Similar response profiles were observed in the intraparietal sulcus (IPS) but the imaging literature implicates the IPS in attentional processing (Corbetta and Shulman, 2002), which could account for this activity as the more beautiful stimuli would also be the most engaging. However as the inferior transverse frontopolar gyrus has never been implicated as an attentional area it can be considered the authentic neural correlate of considered, aesthetic beauty. This explanation receives support from a previous study in which a medial frontopolar activation was implicated in the processing of consonant versus dissonant acoustic stimulation, although this activation was slightly dorsal to the inferior transverse frontopolar gyrus activation and was in the right as opposed to left hemisphere (Blood et al., 1999). Experiment 4 is the first to provide evidence that the

medial inferior transverse frontopolar gyrus becomes more active during exposure to beautiful than less beautiful stimuli that are perceived through either the visual or auditory sensory modalities. Furthermore, it is also the first study to demonstrate parametric modulation of response magnitude that is positively correlated with subjective aesthetic ratings at this neuroanatomical locus.

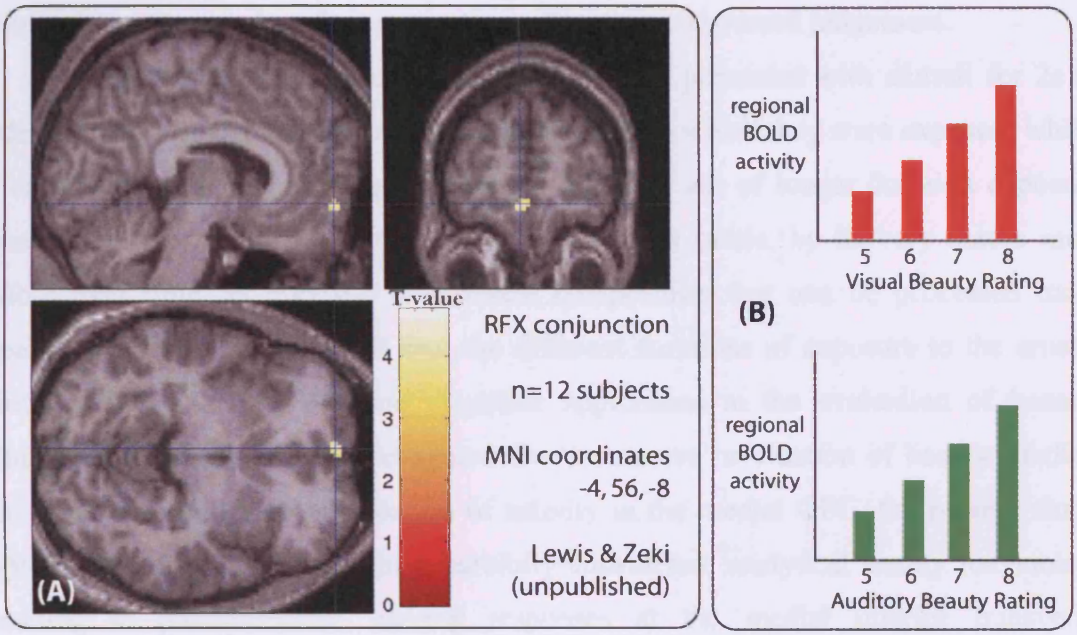


Figure 41: The supramodal beauty area. The conjunction of visual and auditory 1st order parametric contrasts revealed a common area of the inferior transverse frontopolar gyrus overlaid on structural sections (A), with parametrically modulated response profiles that demonstrated a positive linear correlation with the subjective aesthetic rating given to each piece (B, illustrative figures only).

The final conclusion of the Nakamura study (Nakamura et al., 1998) stated that, "left frontal regions are involved in the assessment of facial attractiveness". The findings of experiment 4 and those of the previous fMRI study of visual beauty (Kawabata and Zeki, 2004) both support the notion of left lateralisation of frontal brain activations involved in the judgement of beauty. However, despite the agreement regarding the involvement of left-lateralised OFC in beauty judgements, whilst the Kawabata & Zeki study demonstrated increased activity in the medial OFC in response to beautiful versus neutral or ugly stimuli, this study isolated significant parametrically varying activations to beautiful stimuli in a dorsally adjacent area.

Medial OFC activation has been identified in several imaging experiments in which subjects were required to select response based on instinctive concepts such as

'rightness', 'luck' or 'familiarity' (Elliott et al., 1999a; Elliott et al., 1997; Elliott et al., 1999b; Maguire et al., 1999). The involvement of the medial OFC across such a range of experiments, in which both task and stimuli varied considerably, can be explained by the unifying theme of making responses based on instinct or "... the feeling of 'rightness' rather than any rational analysis... " (Elliott et al., 2000b). This reasoning could be applied to the Kawabata and Zeki study where subjects' rapid appraisal of the painting's beauty was based on instinctive rather than considered judgement.

In the Kawabata & Zeki study, subjects were presented with stimuli for 2s in order to maximise the number of different paintings to which they were exposed, whilst in experiment 4, stimuli were presented for 12s. The use of longer duration exposure time was in fact primarily motivated by the fact that music, by its very nature must unfold over time, as opposed to a visual composition that can be processed more rapidly. It is therefore proposed that the different durations of exposure to the artistic stimuli resulted in two different cognitive approaches to the evaluation of beauty. Whilst the Kawabata study investigated the 'instinctive' evaluation of beauty, finding that it is associated with modulation of activity in the medial OFC, the present study captured the neural correlates of a carefully considered, analytical beauty judgement, resulting in parametrically varying responses at the medial inferior transverse frontopolar gyrus. The differences between this study and this laboratory's previous fMRI beauty study can therefore be accounted for in terms of the two different strategies used to make the beauty judgement.

4.6 Proposed functional specialisation of the right anterior frontal operculum

4.6.1 Breaking the electrophysiological laws of integration

The observation of greater activity under asynchronous than synchronous conditions at the right anterior insula and anterior frontal operculum (aFO) is in general agreement with the findings of the Bushara group, but directly contravenes that which would be predicted by application of the electrophysiological laws of crossmodal integration (Calvert et al., 2001; Stein, 1993). These laws presuppose response depression under asynchronous conditions and response enhancement under synchronous conditions; the implication being that synchronous crossmodal input should always induce greater activity than asynchronous crossmodal input. For asynchronous bimodal stimuli to induce greater activation than the same stimuli under synchronous conditions, this region must either not be a site of crossmodal integration, or these laws may not always be applicable to sensory processing of complex, visual-auditory stimuli in the human neocortex. Indeed, in experiment 2, the CMI conjunction analysis did not identify significant interactions between visual and auditory stimulus processing in the right aFO for synchronous or asynchronous pairs and so the mechanism by which the increased activation in the aFO arises under conditions of bimodal asynchrony remains a mystery. However, that some region or other in the brain must be sensitive to deviations from a correlation between sensory signals in order for accurate perceptual grouping to be achieved, was predicted in a recent review of the psychophysical evidence relating to Bayesian integration strategies that might permit robust perception:

> " ...sensory signals... are most likely to be integrated if they occur simultaneously with no spatial discrepancy, and are not likely to be integrated if the spatial discrepancy is large, or if the temporal sequence of events is not appropriate. That is, with large discrepancies, robust behaviour might be observed in which a discrepant source is discounted or 'vetoed'."
>
> (Ernst and Bulthoff, 2004)

Use of the word 'vetoed' here and discussion of multisensory integration being 'permitted' in the conclusion of the Bushara study (Bushara et al., 2001), suggests a system of regulation for the integration of separate component cues derived from a single perceptual object. Such governance of crossmodal integration may be realised in the right-lateralised anterior insula-aFO complex.

4.6.2 A putative role for the right anterior frontal operculum in multisensory grouping

The right anterior Frontal operculum has been named as such, based on the neuroanatomical location of the reproducible crossmodal asynchrony-sensitive functional activation within the anterior-most part of the frontal operculum, i.e. the part of the pars orbitalis (BA47) that directly overlies the anterior most of the insular gyri - the anterior short insular gyrus (figure 42).

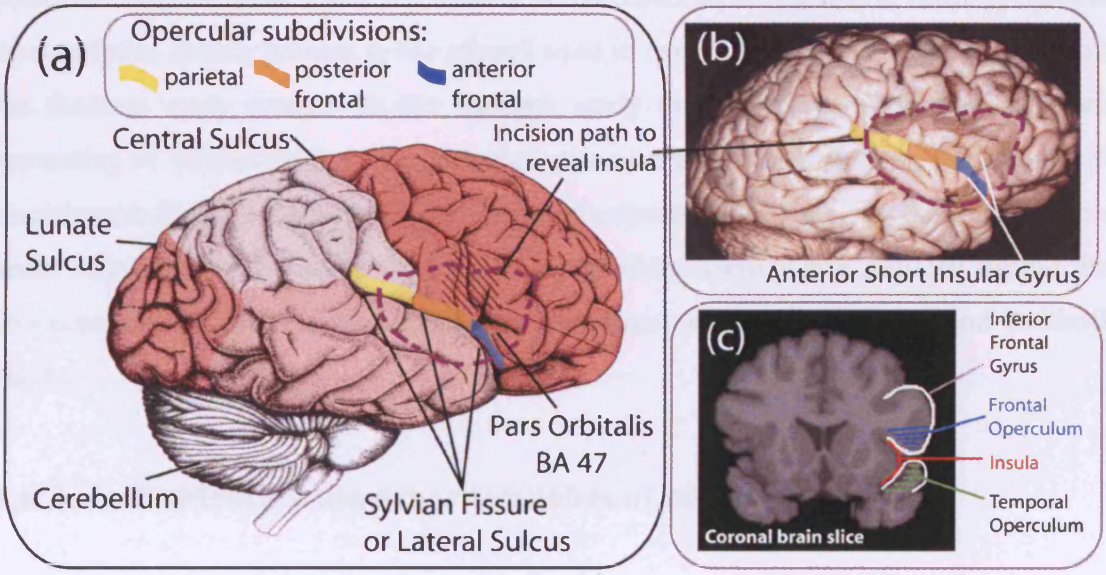


Figure 42: Anatomical description of the anterior frontal operculum. A drawing of the lateral surface of the brain (a), photograph of the lateral view of a human brain with the frontal, parietal and temporal operculi removed to reveal the insula (b) and coronal section through a structural MRI brain image (c) are provided to aid localisation of the crossmodal asynchrony-sensitive region within the right anterior frontal operculum (aFO). The anterior frontal operculum is a subdivision of the posterior pars orbitalis section of the inferior frontal gyrus (BA 47), buried deep in the Sylvian fissure, which overlies the anterior short insular gyrus.

If monitoring deviation from covariation across different sensory streams for crossmodal perceptual grouping is the functional specialisation of the aFO, then it would be expected to be specifically activated only by crossmodal asynchrony and not intramodal asynchrony. This was demonstrated in experiment 3, where significant activation of the right aFO was observed in the contrast of crossmodal asynchrony versus synchrony, but not intramodal asynchrony versus synchrony. Having established that the right aFO is specifically sensitive to crossmodal asynchrony, an explanation is needed for why it was not implicated in the Bushara study (Bushara et al., 2001).

The right superficial aFO may operate in tandem with the deeper anterior-most part of the right insula in the following way: if the activity in the right anterior insula is correlated with the detection of individual, discrete, asynchronous crossmodal events, the activity at the overlying right aFO may be sensitive to the asynchrony between complex, dynamic visual and auditory cycles. This conclusion was reached through the observation that anterior insular activations were detected in experiment 2, experiment 3 and the Bushara study (Bushara et al., 2001), whilst the right aFO activations were only detected in experiments 2 and 3. Thus the involvement of the right aFO must be induced by a stimulus quality present in the stimuli used in experiments 2 and 3, but absent from the Bushara study stimuli. In the Bushara study the stimuli were simple and static consisting of 100ms flashes of a coloured disc and brief auditory tone, whilst in both experiments 2 and 3 the stimuli were complex, dynamic and cyclical. Activity in these areas may collectively 'veto' inappropriate integration of visual and auditory stimuli that co-occur, but do not co-vary, as suggested by Ernst and Bulthoff (Ernst and Bulthoff, 2004).

4.6.3 Evidence from other branches of neuroscience

Support from electrophysiological studies

Electrophysiological evidence to support a role for the presence of visual-auditory bimodal ventrolateral prefrontal neurons that were sensitive to crossmodal stimulus onset asynchrony arose from studies in both cat (Loe and Benevento, 1969) and later the rhesus monkey (Benevento et al., 1977). Neuronal interactions were recorded during combined visual and auditory stimulation with responses varying according to the relative timing of the two stimuli. These included cells whose activities

were modulated as a function of crossmodal stimulus onset asynchrony, i.e. response depression when visual stimuli preceded auditory stimuli by 25ms and response enhancement when the auditory stimuli followed visual stimuli after an interval of 50ms (Benevento et al., 1977). Therefore a population of bimodal neurons in the ventrolateral prefrontal cortex were already known to be highly sensitive to visual-auditory asynchrony in the primate brain. Furthermore, it is known that the reversible cryogenic deactivation of the cat anterior ectosylvian sulcus / rostral lateral sulcus leads to abolition of CMI responses at deep layer superior colliculus neurons (Wallace and Stein, 2000). If the assumption can be made that the aFO is the human homologue of these cortical areas, it would suggest that the sensitivity of the aFO to crossmodal synchronicity may be integrated in some way with that of the superior colliculus.

Support from anatomical connectivity studies

Recent comparison between human and monkey ventrolateral frontal cortex (Petrides and Pandya, 2002) has established that the anatomical location of Brodman area 47 in the human brain (of which the aFO is a part) is architectonically equivalent to the monkey area 12 (Walker, 1940). Injection of retrograde tracers into area 12 revealed inputs from both high level visual (primarily rostral inferotemporal cortex) and also auditory (anterior superior temporal sulcus) sensory areas (Petrides and Pandya, 2002). Therefore, in addition to the tecto-thalamo-insular pathway proposed in the Bushara study as the neural substrate for detecting visual-auditory onset mismatches at the anterior insula (Bushara et al., 2001), the inputs from high level visual and auditory sensory areas might provide the sensory input required to perform asynchrony judgements with more complex, dynamic stimuli. The anatomical connections providing the appropriate level of sensory information to the right aFO to support its putative role in regulating multisensory grouping between complex dynamic stimuli are therefore already established.

Support from the neurodevelopmental literature

The proposal that the aFO may be intrinsically involved in regulation of sensory perception also converges with evidence from the human neurodevelopmental studies of Paul Flechsig (Flechsig, 1901). There is direct anatomical overlap between Flechsig's area 14 and the aFO (figure 43), which is categorised as one of the 'intermediate' areas that receives preferential myelination soon after birth; a privilege shared by other

important sensory areas such as V5 (Watson et al., 1993). Presumably the rapid transmission of action potentials, made possible by myelination due to the saltatory conduction that it permits, is vitally important for V5 to perform its computations regarding visual motion processing. In light of the role proposed here for the aFO - namely the adjudication of appropriate bimodal stimuli for crossmodal integration according to temporal covariation cues - there would clearly be an equivalent pressure on the development of rapid action potential propagation for this structure. In order for the right aFO to perform its putative role in crossmodal asynchrony detection, it may well require myelinated connections to other areas. It would presumably need to receive all relevant sensory information and to send an efferent 'veto' signal as quickly as possible in order to be effective in preventing erroneous crossmodal integration between concurrent but non-covarying stimuli. This time pressure would be even greater if the afferent input to this area does indeed arise primarily from unimodal domains of the rostral temporal cortex, which would indicate that the sensory information it receives must already have undergone considerable (time consuming) unimodal sensory processing prior to its arrival at the aFO. Preferential myelination soon after birth in the aFO might thus support a special role in monitoring the synchrony of multisensory inputs, in order to permit crossmodal integration only when appropriate.

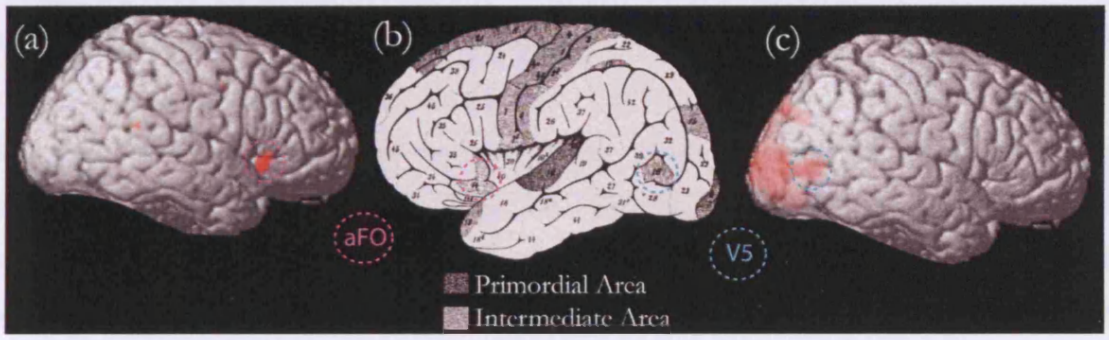


Figure 43: The aFO is a Flechsig 'intermediate' area. The location of the aFO coincides with a subset of regions defined as 'intermediate' by Paul Flechsig based on his studies investigating the chronology of myelinisation in peri-natal human cortical tissue. Primordial areas include the visual and auditory cortex and are fully myelinated at birth. Intermediate areas, such as the human visual motion area V5 and the aFO, receive preferential myelination during the first few post-natal weeks.

Support from the developmental psychology literature

The chronology of myelination development is compatible with developmental behavioural studies of intersensory perception. A popular investigative technique

involves exposing infants to two adjacent visual stimuli with concurrent auditory stimulation that proceeds simultaneously with one, but not the other, visual array. By monitoring the amount of time that infants spend looking at each of the two visual arrays over a period of 30 to 240 seconds, the subjects were observed to spend a significantly greater proportion of time dwelling on the visual array that was temporally synchronised with the auditory stimulus than the other (Spelke, 1976). These studies indicated that infants first demonstrate sensitivity to crossmodal synchronicity from the age of about three months old onwards. However as infants do not appear to be able to discriminate between crossmodally synchronous and asynchronous stimuli in the first few weeks after birth, this appears to tally with the observations made by Flechsig indicating that the myelination of the aFO has not reached full maturity at this stage in brain development.

4.6.4 Temporal predictive error coding for multisensory grouping

Based on the findings of experiment 2 and 3, it is proposed that the aFO complex (i.e. the right aFO and underlying anterior insula) is the neural correlate of dynamic crossmodal asynchrony detection and is specifically sensitive to deviation from temporal covariation between complex, cyclical visual and auditory stimuli. High levels of activity result from exposure to temporally decorrelated crossmodal cues when subjects perceive an asynchronous visual-auditory stimulus array, whilst low activity results during exposure to the synchronous visual-auditory stimulus array - reflecting the tight crossmodal co-variation. Hence, the aFO may operate according to the principles of predictive coding (Rao and Ballard, 1999), but in the context of crossmodal temporal contiguity. In this theoretical framework, a residual error signal would be propagated during deviation from 'expected' crossmodal temporal covariation during concurrent visual and auditory stimulation and this deviation would occur under conditions of crossmodal asynchrony, but not synchrony. Where in the brain this error signal should be propagated to is uncertain, but given the elevated responses to crossmodally synchronous versus asynchronous conditions in the amygdala and the indications of crossmodal integration at the right posterior temporo-parietal junction, it is possible that inhibitory projections may be sent to either or both of these structures.

4.7 Conclusions

The aim of this thesis was to assess where in the human brain extraction of information common to visual and acoustic stimuli occurs and to isolate regions that are sensitive to crossmodal synchronicity, which might therefore regulate visual-auditory integration. This aim was achieved by using fMRI to measure regional brain activations whilst subjects were exposed to a wide variety of visual and auditory stimuli, under various task conditions, which enabled the desired responses to be elicited.

In scanning experiment 1, subjects were required to recognise noise-degraded video footage of everyday occurrences, under bimodal and unimodal conditions, in order to discover where in the brain visual and auditory information becomes crossmodally integrated. These neuroanatomical sites were isolated using the crossmodal integration conjunction, which revealed voxels where statistically significant visual-auditory interactions and bimodal elevations occurred. Bilateral temporo-parietal junction (TPJ) was the only brain region to demonstrate both visual-auditory interactions and an elevation in response under bimodal versus unimodal conditions in this study. The right TPJ cluster was located slightly posterior to the neuroanatomical locus implicated in previous bottom-up attentional studies (e.g. Downar *et al*, 1999) and so was considered to be the neural correlate of crossmodal integration for unified bimodal object perception.

In experiments 2 and 3, temporal synchrony between complex, cyclical visual and acoustic stimuli was manipulated to determine which brain areas were differentially responsive to crossmodal synchrony versus asynchrony. Significant differences in regional brain activation to crossmodal synchrony versus asynchrony would be expected of the neural substrate for a system that can distinguish between related and unrelated bimodal stimuli.

Although in experiment 2 no brain area was found to be more active under conditions of crossmodal synchrony than asynchrony, in experiment 3 the amygdala was more active, bilaterally, during crossmodal synchrony than asynchrony. The amygdala is generally accepted to become active in response to behaviourally relevant sensory stimuli (e.g. fearful faces, sudden loud noises etc.). Elevated activity during crossmodal synchrony may reflect increased salience that results from specific pairs of visual and auditory stimuli being perceived as crossmodally correlated and therefore

relating to a single rather than independent stimuli, or could even drive this very process. Such a role for the amygdala in crossmodal perception has never previously been reported in the literature.

In experiment 2 the right anterior frontal operculum (aFO) and underlying anterior insula (the aFO complex) were more active under conditions of crossmodal asynchrony than synchrony. This activation was not significantly modulated by visual submodality (i.e. unaffected by use of colour versus motion stimuli) or regularity (i.e. unaffected by using cycle periods of regular versus irregular length). The involvement of the right anterior insula in crossmodal asynchrony detection was already known from previous experiments (Bushara et al., 2001), but the involvement of the right aFO is a novel finding. In experiment 3 this finding was confirmed and refined by demonstrating that the right aFO was exclusively sensitive to crossmodal asynchrony, as no significant voxels were detected at this locus in the contrast of intramodal asynchrony versus synchrony. These findings converge with peri-natal neurodevelopmental studies focusing on central neuronal myelination (Flechsig, 1901) and developmental psychology studies investigating the multisensory discriminative abilities of infants (Spelke, 1976). Taken together these studies consolidate the observation that the right aFO is sensitive to crossmodal synchronicity. They suggest that the myelination of aFO neurons only reaches full maturity during post-natal development (Flechsig, 1901) and that the ability to discriminate between crossmodally synchronous and asynchronous stimuli also develops during the first few post-natal weeks of sensory development (Spelke, 1976).

In experiment 4, subjects were asked to evaluate the beauty of stimuli conveyed either through the visual or auditory sensory modalities, or both. In using this approach, which required subjects to extract a common feature (i.e. subjective beauty) from either visual or acoustic stimuli, the brain areas involved in evaluating a single metric irrespective of sensory modality could be revealed. A region of the medial frontal pole cortex was isolated at which the response magnitude was parametrically modulated according to each subject's own aesthetic ratings for visual art and for acoustic musical extracts. It is therefore proposed that the neural correlates of supramodal aesthetic evaluation are located at the medial inferior transverse frontopolar gyrus and that this is a site of multisensory areal convergence (Meredith, 2002). This inference was based on the graded response at this neuroanatomical locus that was positively correlated with the

degree of subjective stimulus beauty, irrespective of the sensory modality through which it was detected.

To summarise, this thesis has provided insights into the neuroanatomical sites at which information is extracted from visual and auditory stimuli beyond the level of unimodal sensory processing. Specifically, the neural correlates of supramodal aesthetic evaluation have been identified in the inferior transverse frontal polar gyrus, as has a possible site at which visual and auditory information might become crossmodally integrated in unified bimodal perception at the right pTPJ. On the basis of further empirical observations, a novel role for the amygdala in detecting crossmodal temporal correlations has been proposed. Finally, it has been proposed for the first time that the right aFO is sensitive to dynamic crossmodal asynchrony and that the resulting neuronal activity in this region may play a role in preventing inappropriate perceptual unification of concurrent, but asynchronously covarying, visual and auditory stimuli.

Part 5:

APPENDICES

5.1 Appendix for experiment 1

5.1.1 Description of film clips

01) 'aft' - applying aftershave
Description: scene shot face-on, actor's head and torso in view, actor sprays aftershave into a hand, hands rubbed together, vigorously applied to face, and then both cheeks slapped twice.

02) 'air' - airhockey
Description: a two player game played on a table with multiple upward facing air jets minimising the friction between table and puck. When hit towards one or other goal the puck moves quickly and bounces noisily off the sides of the table. Shot from side view, one player hits the puck which ricochets off the side walls three times before the other player hits it back. The puck then bounces into the goal off the adjacent wall.

03) 'bad' - badminton
Description: a game played with two rackets and a feathered shuttlecock hit over a head-height net in the middle of the court. Camera angle was from the vantage point of an elevated balcony level with the back of the court with 2 participants playing a six shots rally.

04) 'bgo' - bongo
Description: a busker standing in a tunnel of the London underground playing the bongo in an animated fashion shot from head to knees.

05) 'bmb' - bomb
Description: shot in a swimming pool from behind an actor who bends down and leaps into the air, tucking into a ball and dropping into the water with a splash.

06) 'ccn' - coin counting
Description: close-up shot of a pair of hands and a money tray with coins dropped one-by-one into the cash register.

07) 'cfl' - connect four
Description: a close up of the popular game where two players take turns dropping yellow and red tokens into one of eight columns to try and get four in a row. Only the hands of the players come into view at the top edge of the clip and otherwise the half full grid of tokens is visible. Players have one turn each and open the bottom of the grid to allow all the tokens to slide out onto the table.

08) 'chp' - scooping up poker chips
Description: close-up of an actors arms and table with poker chips on it. The actor leans forwards (towards camera) extends arms and hands around the pile of chips, drags them toward him and then begins to stack them in piles

09) 'deo' - deodorant
Description: front view of seated actor who raises his arm and applies aerosol deodorant to his underarm in two long then one short burst.

10) 'dog' - scampering dog
Description: shot indoors from 1m above ground with a large dog running towards the camera, turning 90^{o} in mid-stride and then skidding to a halt on the slippery parquet floor.

11) 'scp' - spinning chip
Description: close up of a poker chip spinning on its edge with decreasing rotational velocity until it drops onto its flat side and stops.

12) 'gsw' - garden sweep
Description: shot outside in a garden face on with an actor sweeping dust towards him from the left, centre and right.

13) 'rip' - ripping newspaper
Description: actor seated on a chair holding a broadsheet newspaper open which is then torn along the central crease by pulling the corners away from each other.

14) 'row' - rowing machine
Description: actor is shot from the rear right whilst using rowing machine with strokes of about 0.5Hz.

15) 'rub' - rubbish truck
Description: actor hooks restaurant bin to the dump truck, which inverts the bin and tips the contents inside.

16) 'saw' - hack saw
Description: close up of actor's hands as he saws a piece of wood with a hacksaw using strokes of about 1Hz.

17) 'shf' - card shuffle
Description: close up of actor's hands over a lit table which shuffle playing cards by dividing the pack into two, holding both halves down on the table and using the thumbs to randomly overlap the card edges.

18) 'skp' - skipping rope
Description: full body shot of actor using a skipping rope making jumps at an approximate rate of 2Hz.

19) 'str' - spoon stirring a cup
Description: close up of finger tips holding a spoon which is used to stir a mug of tea.

20) 'tbr' - brushing teeth

Description: front view of seated actor rapidly brushing teeth with circular brushing motion on clenched teeth

5.1.2 Visual white noise script for 'avi' animations

```
screenwidth=240
screenheight=320
frames = 25*6; % (to make 6s animation of 25 frames per second)
for f = 1:frames
  n=0;
    for j=1:screenheight
      n=((j-1)*screenwidth)+1;
      p = j.*screenwidth;
      x(n:p,f)=(1:screenwidth)';
      y(n:p,f)=ones(screenwidth,1).*j;
      c(n:p,f)=round(rand(screenwidth,1));
    end
  disp(num2str(f))
end
config_display(1,1,[0,0,0],[1,1,1],'Arial',50,8,0);
cgloadlib
cgopen(1,8,0,1)
cgcoltab(0,0,0,0);      %black=0
cgcoltab(1,0.5,0.5,0.5); %grey=1
cgcoltab(2,1,1,1);      %white=2
cgnewpal;
rand=randperm(1)
for f = rand
  cgdraw(x(:,f)-screenwidth/2,y(:,f)-screenheight/2,c(:,f))
  cgflip(0)
  cgscrdmp
end
```

5.2 Matlab script for experiment 2

Continuous cycling pure tone script

Cycle_vector = sin(pi*(440 - 220)/T*(t + f1/g).^2)

Where: T is half the cycle period in seconds

t is a vector of numbers from zero to T in intervals of 1/sample rate

g is the rate of pitch change i.e. (440-220)/T

The sign of variable 'g' controlled whether it was the ascending (+), or the descending (-) part of the cycle. Therefore, two vectors, one for each of the ascending and another for the descending half of the cycle were required to create a single pitch cycle or to drive a single visual dot array cycle.

5.3 Appendix for experiment 4

List of art and music stimuli:

5.3.1 VISUAL STIMULI

ID#01	Jackson Pollock	Number 8
ID#02	Marc Chagall	The Good Samaritan
ID#03	Marc Chagall	House
ID#04	Joan Miro	Waggon Tracks
ID#05	Camille Pissaro	Jallais Hill, Pointoise
ID#06	Jan Vermeer	The Glass of Wine
ID#07	Robert Holmes	Deux
ID#08	Jackson Pollock	Oranges and Lemons
ID#09	George Grosz	The city
ID#10	Marcel Duchamp	Young girl & man in spring
ID#11	Michelangelo	Rondanini Pieta
ID#12	Willem DeKooning	Untitled
ID#13	Paul Signac	View of St. Tropez
ID#14	Claude Monet	Woman with an Umbrella
ID#15	Michelangelo	Day, Tomb of Guiliano de Medici
ID#16	Alberto Giacometti	Dog
ID#17	Jackson Pollock	Flame
ID#18	Alberto Giacometti	Standing woman
ID#19	Andre Derain	Fishing Port, Collioure
ID#20	Georges Braque	Castle La Roche-Guyon
ID#21	Willem DeKooning	Untitled
ID#22	Marcel Duchamp	Portrait of chess players
ID#23	Gian Lorenzo Bernini	Daniel and the lion
ID#24	Gian Lorenzo Bernini	The ecstasy of St. Therese
ID#25	Max Ernst	Elephant Celebes
ID#26	Magritte	Dove with flower tail
ID#27	Marcel Duchamp	Nude descending staircase #2
ID#28	Henri Matisse	Interior with a violin case
ID#29	Titian	Venus blindfolding cupid
ID#30	Georges Braque	Fruit Dish, Ace of Clubs
ID#31	Willem de Kooning	Tor zum Fluss
ID#32	Kazimir Malevich	Supremus No. 55
ID#33	Vincent Van Gough	Noon rest
ID#34	Joseph Turner	Alnwick Castle, Northumberland
ID#35	Franz Kline	Untitled
ID#36	Willem DeKooning	Untitled
ID#37	Robert Delauney	Homage to Bleriot
ID#38	Rene Magritte	Bather
ID#39	Barnett Newman	Untitled
ID#40	Jan Vermeer	View of Delfi

5.3.2 ACOUSTIC STIMULI

ID#01	Charles Mingus	Haitian fight song
ID#02	Gustav Mahler	Symphonie No.3: IV, mid-section
ID#03	John Barry	Dances with wolves, late mid-section
ID#04	Unknown	Dixie blues
ID#05	Charles Mingus	Goodbye pork pie hat, intro-section
ID#06	Richard Wagner	Ring cycle, Coda
ID#07	Charles Mingus	Shoes of a fisheman's wife..., mid-section
ID#08	Ludwig Van Beethoven	Late piano sonatas: track 1, intro-section
ID#09	John Barry	Dances with wolves, late mid-section
ID#10	Charles Mingus	Peggy's blue skylight, end section
ID#11	Ludwig Van Beethoven	Spring sonata
ID#12	Charles Mingus	Better git it in your soul, end section
ID#13	Joseph Haydn	String quartet in D, presto
ID#14	Johannes Sebastian Bach	Violin concerto in E, mid-section
ID#15	Fryderyk Chopin	Balade No. 4
ID#16	John Coltrane	Its easy to remember, mid-section
ID#17	Wolfgang Mozart	Oboe quartet in F major
ID#18	Fryderyk Chopin	Mazurkas op.59
ID#19	John Coltrane	Too young to go steady, intro-section
ID#20	Miles Davis	Old folks, end-section
ID#21	Ludwig Van Beethoven	Wind concerto, track 9
ID#22	Charles Mingus	Shoes of a fisheman's wife..., end section
ID#23	Ludwig Van Beethoven	Piano trio: track 1, intro-section
ID#24	Miles Davis	Old folks, intro-section
ID#25	Ludwig Van Beethoven	Late piano sonatas: track 1, mid-section
ID#26	Franz Schubert	Winterreise: Gute Nacht
ID#27	Gustav Mahler	Symphonie No.3: IV, intro
ID#28	Unknown	Farewell and end title
ID#29	Miles Davis	My ship, intro-section
ID#30	John Williams	Saving private ryan, early mid-section
ID#31	Miles Davis	Summer night, intro-section
ID#32	Miles Davis	Time after time, mid-section
ID#33	Wolfgang Mozart	Fantasie in C minor
ID#34	John Coltrane	You don't know what love is
ID#'35	Duke Ellington	The jeep is jumpin', mid-section
ID#36	Wes Montgomery	Here's that rainy day, mid-section
ID#37	Miles Davis	I fall in love too easily, intro-section
ID#38	Charles Mingus	Peggy's blue skylight, intro-section
ID#39	Pharoh Saunders	Thembi, intro-section
ID#40	Ludwig Van Beethoven	Wind concerto, track 6, end-section

Part 6:

REFERENCES

REFERENCES

Adamuk, D. (1870). Uber die Innervation der Augenbewegungen. Zentralbl Med Wiss *8*, 65-67.

Alais, D., and Burr, D. (2003). The "Flash-Lag" effect occurs in audition and cross-modally. Curr Biol *13*, 59-63.

Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr Biol *14*, 257-262.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends Cogn Sci *4*, 267-278.

Altman, J. A., Balonov, L. J., and Deglin, V. L. (1979). Effects of unilateral disorder of the brain hemisphere function in man on directional hearing. Neuropsychologia *17*, 295-301.

Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992). Anatomical organisation of the primate amygdaloid complex. In The Amygdala: Neurobiological Aspects, of Emotion , Memory and Mental Dysfunction, J. P. Aggleton, ed. (New York, Wiley-Liss), pp. 1-66.

Amassian, V. E., and Devito, R. V. (1954). Unit activity in reticular formation and nearby structures. J Neurophysiol *17*, 575-603.

Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. (1997). Multimodal integration for the representation of space in the posterior parietal cortex Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Philos Trans R Soc Lond B Biol Sci *352*, 1421-1428.

Anderson, F. D., and Berry, C. M. (1959). Degeneration studies of long ascending fiber systems in the cat brain stem. J Comp Neurol *111*, 195-229.

Apter, J. T. (1945). Projection of the retina on the superior colliculus of the cats. Journal of Neurophysiology *9*, 73-86.

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., and Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. J Neurosci *23*, 9632-9638.

Aschersleben, G., and Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. Int J Psychophysiol *50*, 157-163.

Azizi, S. A., and Woodward, D. J. (1990). Interactions of visual and auditory mossy fiber inputs in the paraflocculus of the rat: a gating action of multimodal inputs. Brain Res *533*, 255-262.

Bar, M., and Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. Proc Natl Acad Sci U S A *96*, 1790-1793.

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., and Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. Neuron *29*, 529-535.

Barlow, H. B. (1986). Why have multiple cortical areas? Vision Res *26*, 81-90.

Baron-Cohen, S., Harrison, J., Goldstein, L. H., and Wyke, M. (1993). Coloured speech perception: is synaesthesia what happens when modularity breaks down? Perception *22*, 419-426.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., and Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci *17*, 377-391.

Bartels, A., and Zeki, S. (1998). The theory of multistage integration in the visual brain. Proc R Soc Lond B Biol Sci *265*, 2327-2332.

Bartels, A., and Zeki, S. (2000). The architecture of the colour centre in the human visual brain: new results and a review. Eur J Neurosci *12*, 172-193.

Battaglia, P. W., Jacobs, R. A., and Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am A Opt Image Sci Vis *20*, 1391-1397.

Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Henaff, M. A., Michel, F., and Barton, J. J. (2001). Unilateral right parietal damage leads to bilateral deficit for high-level motion. Neuron *32*, 985-995.

Battersby, W. S., Bender, M. B., Pollack, M., and Kahn, R. L. (1956). Unilateral spatial agnosia (inattention) in patients with cerebral lesions. Brain *79*, 68-93.

Baylis, G. C., Rolls, E. T., and Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. J Neurosci *7*, 330-342.

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol *15*, 145-153.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004a). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci *7*, 1190-1192. Epub 2004 Oct 1110.

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron *41*, 809-823.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., and Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. Neuroimage *14*, 310-321.

Bechara, A., Tranel, D., Damasio, H., and Damasio, A. R. (1996). Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb Cortex *6*, 215-225.

Bell, C., Sierra, G., Buendia, N., and Segundo, J. P. (1964). Sensory Properties of Neurons in the Mesencephalic Reticular Formation. J Neurophysiol *27*, 961-987.

Bellmann, A., Meuli, R., and Clarke, S. (2001). Two types of auditory neglect. Brain *124*, 676-687.

Benevento, L. A., Fallon, J., Davis, B. J., and Rezak, M. (1977). Auditory--visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol *57*, 849-872.

Benevento, L. A., and Fallon, J. H. (1975). The ascending projections of the superior colliculus in the rhesus monkey (Macaca mulatta). J Comp Neurol *160*, 339-361.

Bernstein, I. H., Clark, M. H., and Edelstein, B. A. (1969). Effects of an auditory signal on visual reaction time. J Exp Psychol *80*, 567-569.

Bernstein, I. H., Rose, R., and Ashe, V. M. (1970). Energy integration in intersensory facilitation. J Exp Psychol *86*, 196-203.

Berridge, K. C. (1996). Food reward: brain substrates of wanting and liking. Neurosci Biobehav Rev *20*, 1-25.

Bertelson, P., and Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. Int J Psychophysiol *50*, 147-155.

Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept Psychophys *29*, 578-584.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex *10*, 512-528.

Bisiach, E., Cornacchia, L., Sterzi, R., and Vallar, G. (1984). Disorders of perceived auditory lateralization after lesions of the right hemisphere. Brain *107*, 37-52.

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci U S A *98*, 11818-11823.

Blood, A. J., Zatorre, R. J., Bermudez, P., and Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nat Neurosci *2*, 382-387.

Breiter, H. C., and Rosen, B. R. (1999). Functional magnetic resonance imaging of brain reward circuitry in the human. Ann N Y Acad Sci *877*, 523-547.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K., and Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. Neuron *29*, 287-296.

Broadbent, D. E., and Broadbent, M. H. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. Percept Psychophys *42*, 105-113.

Broadbent, J. M. (1989). Essence of the beautiful face. Funct Orthod *6*, 18-22, 24.

Broca, P. P. (1861). Perte de la parole, ramollisement chronique et destruction pertielle du lobe anterieure gauche de cerveau. Bull Soc Anthropol Paris *2*, 235-238.

Brodman, K. (1905). Beitraege zur histologischen Lokalisation der Grosshirnrinde Dritte Mitteilung: Die Rindenfelder der niederen Affen. J Psychol Neurol Lpz *47*.

Brodman, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde (Liepzig, Barth).

Brown, S., and Schafer, A. (1888). An investigation into the functions of the occipital and temporal lobes of the monkey's brain. Philos Trans R Soc Lond B Biol Sci *179*, 303-327.

Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol *46*, 369-384.

Buchel, C., Morris, J., Dolan, R. J., and Friston, K. J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. Neuron *20*, 947-957.

Buchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., and Friston, K. J. (1996). Nonlinear regression in parametric activation studies. Neuroimage *4*, 60-66.

Bushara, K. O., Grafman, J., and Hallett, M. (2001). Neural Correlates of Auditory-Visual Stimulus Onset Asynchrony Detection. J Neur *21*, 300-304.

Caan, W., Delgado-Garcia, J., Stein, J. F., and Wattam-Bell, J. (1976). Proceedings: Interaction of visual and auditory inputs to cerebellar Purkinje cells in cat posterior vermis. J Physiol *258*, 20P-21P.

Callan, D. E., Callan, A. M., Kroos, C., and Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Brain Res Cogn Brain Res *10*, 349-353.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex *11*, 1110-1123.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., and David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport *10*, 2619-2623.

Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol *10*, 649-657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage *14*, 427-438.

Celesia, G. G. (1976). Organization of auditory cortical areas in man. Brain *99*, 403-414.

Chalupa, L. M., and Fish, S. E. (1978). Response characteristics of visual and extravisual neurons in the pulvinar and lateral posterior nuclei of the cat. Exp Neurol *61*, 96-120.

Chapman, C. E., Spidalieri, G., and Lamarre, Y. (1986). Activity of dentate neurons during arm movements triggered by visual, auditory, and somesthetic stimuli in the monkey. J Neurophysiol *55*, 203-226.

Chavis, D. A., and Pandya, D. N. (1976). Further observations on corticofrontal connections in the rhesus monkey. Brain Res *117*, 369-386.

Cohen, L., and Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. Neuroimage *22*, 466-476.

Cole, M., Chorover, S. L., and Ettlinger, G. (1961). Crossmodal-transfer in man. Nature *191*, 1225-1226.

Colquhoun, W. P. (1975). Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. Hum Factors *17*, 425-437.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci *3*, 292-297.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci *11*, 2383-2402.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci *3*, 201-215.

Cragg, B. G. (1969). The topography of the afferent projections in the circumstriate visual cortex of the monkey studied by the Nauta method. Vision Res *9*, 733-747.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. Nat Neurosci *7*, 189-195. Epub 2004 Jan 2018.

Cusack, R., Carlyon, R. P., and Robertson, I. H. (2001). Auditory midline and spatial discrimination in patients with unilateral neglect. Cortex *37*, 706-709.

Denny-Brown, D., Meyer, J. S., and Horenstein, S. (1952). The significance of perceptual rivalry resulting from parietal lesion. Brain *75*, 433-471.

Deouell, L. Y., and Soroker, N. (2000). What is extinguished in auditory extinction? Neuroreport *11*, 3059-3062.

Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci *4*, 2051-2062.

Desimone, R., and Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. Brain Res *178*, 363-380.

Diamond, S. P., and Bender, M. B. (1965). On auditory extinction and alloacusis. Trans Am Neurol Assoc *90*, 154-157.

Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. Neuroimage *23*, 64-74.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. Nat Neurosci *3*, 277-283.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. Neuroimage *14*, 1256-1267.

Downar, J., Crawley, A. P., Mikulis, D. J., and Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. J Neurophysiol *87*, 615-620.

Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470-2473.

Driver, J., and Mattingley, J. B. (1998). Parietal neglect and visual awareness. Nat Neurosci *1*, 17-22.

Dubner, R., and Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res *35*, 528-532.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., and Shibuya, H. (1999). Systematic analysis of deficits in visual attention. J Exp Psychol Gen *128*, 450-478.

Duncan, J., Ward, R., and Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. Nature *369*, 313-315.

Duvernoy, H. (1991). The Human Brain: Surface, Three-Dimensional Sectional Anatomy and MRI (New York, Springer-Verlag).

Elliott, R., Dolan, R. J., Frith, C. D., and Rees, G. (2000a). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. Ventromedial prefrontal cortex mediates guessing Differential neural response to positive and negative feedback in planning and guessing tasks. Cereb Cortex *10*, 308-317.

Elliott, R., Dolan, R. J., Rees, G., and Frith, C. D. (1999a). Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals Ventromedial prefrontal cortex mediates guessing. Differential neural response to positive and negative feedback in planning and guessing tasks. J Neurosci *19*, 5066-5073.

Elliott, R., Friston, K. J., Dolan, R. J., Frith, C. D., and Rees, G. (2000b). Dissociable neural responses in human reward systems. Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. Ventromedial prefrontal cortex mediates guessing. Differential

neural response to positive and negative feedback in planning and guessing tasks. J Neurosci *20*, 6159-6165.

Elliott, R., Frith, C. D., and Dolan, R. J. (1997). Differential neural response to positive and negative feedback in planning and guessing tasks. Neuropsychologia *35*, 1395-1404.

Elliott, R., Rees, G., Dolan, R. J., and Frith, C. D. (1999b). Ventromedial prefrontal cortex mediates guessing. Differential neural response to positive and negative feedback in planning and guessing tasks. Neuropsychologia *37*, 403-411.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598-601.

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429-433.

Ernst, M. O., Banks, M. S., and Bulthoff, H. H. (2000). Touch can change visual slant perception. Nat Neurosci *3*, 69-73.

Ernst, M. O., and Bulthoff, H. H. (2004). Merging the senses into a robust percept. Trends Cogn Sci *8*, 162-169.

Essen, D. C., and Zeki, S. M. (1978). The topographic organization of rhesus monkey prestriate cortex. J Physiol *277*, 193-226.

Everitt, B. J., Morris, K. A., O'Brien, A., and Robbins, T. W. (1991). The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic-striatal interactions underlying reward-related processes. Neuroscience *42*, 1-18.

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci *22*, 5749-5759.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex *1*, 1-47.

Fendrich, R., and Corballis, P. M. (2001). The temporal cross-capture of audition and vision. Percept Psychophys *63*, 719-725.

Ferrier, D. (1876). The Functions of the Brain (London, Smith Elder).

Fitzpatrick, K. A., and Imig, T. J. (1980). Auditory cortico-cortical connections in the owl monkey. J Comp Neurol *192*, 589-610.

Flechsig, P. (1901). Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. Lancet *ii*, 1027-1029.

Flourens, P. (1842). Recherches Expérimentales sur les Propriétés et les Fonctions du Systeme Nerveux dans les Animaux Vertébrés (Paris, J.-B. Bailliere).

Fraisse, P. (1981). Multisensory aspects of rhythm. In Intersensory perception and sensory integration, R. D. Walk, and H. L. J. Pick, eds. (New York, Plenum Press), pp. 217-247.

Freeman, J. A. (1970). Responses of cat cerebellar Purkinje cells to convergent inputs from cerebral cortex and peripheral sensory systems. J Neurophysiol *33*, 697-712.

Friston, K., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp *2*, 189-210.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. (1991). Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab *11*, 690-699.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., and Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. Neuroimage *10*, 385-396.

Fritsch, G. and Hitzig, E. (1870). Uber die elektrische Erregbakeit des Grosshirns. Arch f Anat Physiol u wiss Med *37*, 300-332.

Galaburda, A., and Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. J Comp Neurol *190*, 597-610.

Gall, F. J. and Spurtzheim, J. G. (1809). Recherches sur le system nerveux en general, et sur celui du cerveau en particulier (Paris).

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia *38*, 11-21.

Gallant, J. L., Braun, J., and Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science *259*, 100-103.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain *119*, 593-609.

Galton, F. (1883). Inquiries into human faculty and its development (London, Macmillan).

Gebhard, J. W., and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. Am J Psychol *72*, 521-529.

Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci *11*, 473-490.

Gibson, J. J. (1966). The senses considered as perceptual systems (Boston, Houghton Mifflin).

Giesbrecht, B., Woldorff, M. G., Song, A. W., and Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. Neuroimage *19*, 496-512.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., and Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. J Neurophysiol *84*, 1588-1598.

Gloor, P., Olivier, A., Quesney, L. F., Andermann, F., and Horowitz, S. (1982). The role of the limbic system in experiential phenomena of temporal lobe epilepsy. Ann Neurol *12*, 129-144.

Gordon, B. (1973). Receptive fields in deep layers of cat superior colliculus. J Neurophysiol *36*, 157-178.

Gottfried, J. A., and Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. Nat Neurosci *7*, 1144-1152. Epub 2004 Sep 1112.

Goulet, S., and Murray, E. A. (2001). Neural substrates of crossmodal association memory in monkeys: the amygdala versus the anterior rhinal cortex. Behav Neurosci *115*, 271-284.

Grant, S. J., Aston-Jones, G., and Redmond, D. E., Jr. (1988). Responses of primate locus coeruleus neurons to simple and complex sensory stimuli. Brain Res Bull *21*, 401-410.

Gratiolet, L.-P. (1861). Reprise de la discussion sur le volume et la forme du cerveau. Bull Soc Anthropol *2*, 239-275.

Green, D. M. and Swets, J. A. (1979). Signal detection theory and psychophysics (New York, Wiley).

Greene, J., and Haidt, J. (2002). How (and where) does moral judgment work? Trends Cogn Sci *6*, 517-523.

Griffiths, T. D. (1999). Human complex sound analysis. Clin Sci (Lond) *96*, 231-234.

Griffiths, T. D., Buchel, C., Frackowiak, R. S., and Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. Nat Neurosci *1*, 422-427.

Griffiths, T. D., Penhune, V., Peretz, I., Dean, J. L., Patterson, R. D., and Green, G. G. (2000). Frontal processing and auditory perception. Neuroreport *11*, 919-922.

Griffiths, T. D., Rees, A., Witton, C., Cross, P. M., Shakir, R. A., and Green, G. G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. Brain *120*, 785-794.

Griffiths, T. D., Warren, J. D., Scott, S. K., Nelken, I., and King, A. J. (2004). Cortical processing of complex sound: a way forward? Trends Neurosci *27*, 181-185.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187-203.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. Neuron *21*, 191-202.

Grill-Spector, K., and Malach, R. (2004). The human visual cortex. Annu Rev Neurosci *27*, 649-677.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol *35*, 96-111.

Gross, C. G., Schiller, P. H., Wells, C., and Gerstein, G. L. (1967). Single-unit activity in temporal association cortex of the monkey. J Neurophysiol *30*, 833-843.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., and Blake, R. (2000). Brain areas involved in perception of biological motion. J Cogn Neurosci *12*, 711-720.

Grossman, E. D., and Blake, R. (2002). Brain Areas Active during Visual Perception of Biological Motion. Neuron *35*, 1167-1175.

Guski, R., and Troje, N. F. (2003). Audiovisual phenomenal causality. Percept Psychophys *65*, 789-800.

Hall, D. A. (2002). Spectral and Temporal Processing in Human Auditory Cortex. Cerebral Cortex *12*, 140-149.

Hall, D. A. (2003). Auditory Pathways: Are 'What' and 'Where' Appropriate? Current Biology *13*, R406-408.

Hart, H. C., Palmer, A. R., and Hall, D. A. (2003). Amplitude and Frequency-modulated Stimuli Activate Common Regions of Human Auditory Cortex. Cereb Cortex *13*, 773-781.

Heilman, K. M., Bowers, D., and Watson, R. T. (1983). Performance on hemispatial pointing task by patients with neglect syndrome. Neurology *33*, 661-664.

Heilman, K. M., and Valenstein, E. (1972). Auditory neglect in man. Arch Neurol *26*, 32-35.

Henschen, S. E. (1894). Sur les centres optiques cérébraux. Rev Gén Opthal (Paris) *13*, 337-352.

Heron, J., Whitaker, D., and McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. Vision Res *44*, 2875-2884.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. J Exp Psychol *63*, 289-293.

Hikosaka, K., Iwai, E., Saito, H., and Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J Neurophysiol *60*, 1615-1637.

Hikosaka, O., Sakamoto, M., and Usui, S. (1989). Functional properties of monkey caudate neurons. II. Visual and auditory responses. J Neurophysiol *61*, 799-813.

Hikosaka, O., and Wurtz, R. H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. I. Relation of visual and auditory responses to saccades. J Neurophysiol *49*, 1230-1253.

Hirsh, I. J., and Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. J Exp Psychol *62*, 423-432.

Holmes, G. (1945). The Ferrier Lecture: The organization of the visual cortex in man. Proc R Soc Lond B Biol Sci *132*, 348-361.

Hopfinger, J. B., Buonocore, M. H., and Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. Nat Neurosci *3*, 284-291.

Horn, G., and Hill, R. M. (1966). Responsiveness to sensory stimulation of units in the superior colliculus and subjacent tectotegmental regions of the rabbit. Exp Neurol *14*, 199-223.

Howard, I. P. and Templeton, W. B. (1966). Human Spatial Orientation (NY, Wiley).

Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., and Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. Curr Biol *6*, 1015-1019.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol *160*, 106-154.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. J Physiol *195*, 215-243.

Husain, M., and Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. J Neurol *243*, 652-657.

Husain, M., and Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. Nat Rev Neurosci *4*, 26-36.

Husain, M., Shapiro, K., Martin, J., and Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial neglect patients. Nature *385*, 154-156.

Itaya, S. K., and Van Hoesen, G. W. (1982). Retinal innervation of the inferior colliculus in rat and monkey. Brain Res *233*, 45-52.

Ito, S. I. (1982). Prefrontal unit activity of macaque monkeys during auditory and visual reaction time tasks. Brain Res *247*, 39-47.

Jerison, H. J. (1970). Brain evolution: new light on old principles. Science *170*, 1224-1225.

Jones, B. (1981). The developmental significance of cross-modal matching. In Intersensory perception and sensory integration, R. D. Walk, and H. L. J. Pick, eds. (New York, Plenum Press), pp. 109-132.

Jones, E. G., and Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain *93*, 793-820.

Kaada, B. R. (1972). Stimulation and regional ablation of the amygdaloid complex with reference to functional representations. In The neurobiology of the amygdala, B. E. Eleftheriou, ed. (New York, Plenium Press), pp. 205-282.

Kant, I. (1787). Critique of pure reason. (Translation by Guyer, P. and Wood, A. 1998 Cambridge, Cambridge University Press).

Kanwisher, N., Chun, M. M., McDermott, J., and Ledden, P. J. (1996). Functional imagining of human visual recognition. Brain Res Cogn Brain Res *5*, 55-67.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci *17*, 4302-4311.

Karnath, H. O., Ferber, S., and Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. Nature *411*, 950-953.

Kastner, S., and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. Annu Rev Neurosci *23*, 315-341.

Kawabata, H., and Zeki, S. (2004). Neural correlates of beauty. J Neurophysiol *91*, 1699-1705.

Kerkhoff, G., Artinger, F., and Ziegler, W. (1999). Contrasting spatial hearing deficits in hemianopia and spatial neglect. Neuroreport *10*, 3555-3560.

Killcross, S., Robbins, T. W., and Everitt, B. J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature *388*, 377-380.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. Perception & psychophysics *22*, 19-30.

Kinchla, R. A. and Collyer, C. (1974). Detecting a large letter in briefly presented arrays: a confidence rating analysis in terms of a weighted additive effects model. Perception and psychophysics *16*, 117-122.

King, A. J., and Calvert, G. A. (2001). Multisensory integration: perceptual grouping by eye and ear. Curr Biol *11*, R322-325.

King, A. J., and Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res *60*, 492-500.

Klapper, Z. S., and Birch, H. G. (1971). Developmental course of temporal patterning in vision and audition. Percept Mot Skills *32*, 547-555.

Kluever, H., and Bucy, P. C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. Arch Neurol Psychiat *42*, 979-1000.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. Science *297*, 846-848.

Komatsu, H., Ideura, Y., Kaji, S., and Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. J Neurosci *12*, 408-424.

Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. J Neurosci *20*, 3310-3318.

Krekelberg, B. (2003). Sound and vision. Trends Cogn Sci *7*, 277-279.

Kuypers, H. G., Szwarcbart, M. K., Mishkin, M., and Rosvold, H. E. (1965). Occipitotemporal Corticocortical Connections in the Rhesus Monkey. Exp Neurol *11*, 245-262.

LaBar, K. S., LeDoux, J. E., Spencer, D. D., and Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. J Neurosci *15*, 6846-6855.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology *30*, 261-273.

Lashley, K. S. (1946). Mass action in cerebral function. Science *73*, 245-254.

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., and Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp *19*, 213-223.

Lawson, I. R. (1962). Visual-spatial neglect in lesions of the right cerebral hemisphere. A study in recovery. Neurology *12*, 23-33.

Le, T. H., Pardo, J. V., and Hu, X. (1998). 4 T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions. J Neurophysiol *79*, 1535-1548.

LeDoux, J. E. (1995). Emotion: clues from the brain. Annu Rev Psychol *46*, 209-235.

LeDoux, J. E. (1996). The emotional brain (New York, Simon and Schuster).

Leist, A. (2003). What makes bodies beautiful. J Med Philos *28*, 187-219.

Lewald, J., and Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Brain Res Cogn Brain Res *16*, 468-478.

Lewald, J., and Guski, R. (2004). Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. Neurosci Lett *357*, 119-122.

Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. Cereb Cortex *10*, 873-888.

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J Comp Neurol *428*, 112-137.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., and DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. Cereb Cortex *14*, 1008-1021. Epub 2004 May 1027.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. Psychol Bull *126*, 281-308.

Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol *92*, 204-214.

Liegeois-Chauvel, C., Musolino, A., and Chauvel, P. (1991). Localization of the primary auditory area in man. Brain *114*, 139-151.

Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., and Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. Brain *121*, 1853-1867.

Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. Arch Psychiatr Nervenkr *21*, 222-270.

Liu, T., Slotnick, S. D., Serences, J. T., and Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. Cereb Cortex *13*, 1334-1343.

Locke, J. (1959). An essay concerning human understanding (New York, Dover).

Loe, P. R., and Benevento, L. A. (1969). Auditory-visual interaction in single units in the orbito-insular cortex of the cat. Electroencephalogr Clin Neurophysiol *26*, 395-398.

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. Annu Rev Neurosci *19*, 577-621.

Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M. P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C., and Frackowiak, R. S. (1989). The colour centre in the cerebral cortex of man. Nature *340*, 386-389.

Macaluso, E., George, N., Dolan, R., Spence, C., and Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage *21*, 725-732.

MacKay, G. (1888). A discussion on a contribution to the study of hemianopsia, with special reference to acquired colour-blindness. Br Med J *2*, 1033-1037.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage *14*, 802-816.

Maguire, E. A., Frith, C. D., and Morris, R. G. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. Brain *122*, 1839-1850.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., and Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci U S A *92*, 8135-8139.

Malkova, L., Gaffan, D., and Murray, E. A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but

interfere with reinforcer devaluation effects in rhesus monkeys. J Neurosci *17*, 6011-6020.

Manley, J. A., and Mueller-Preuss, P. (1978). Response variability in the mammalian auditory cortex: an objection to feature detection? Fed Proc *37*, 2355-2359.

Marburg, O., and Warner, J. F. (1947). The pathways of the tectum (anterior colliculus) of the midbrain in cats. Journal of Nervous and Mental Disorders *106*, 415-446.

Marshall, J. C., and Fink, G. R. (2001). Spatial cognition: where we were and where we are. Neuroimage *14*, S2-7.

Massaro, D. W., Cohen, M. M., and Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. J Acoust Soc Am *100*, 1777-1786.

McClure, S. M., York, M. K., and Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of FMRI. Neuroscientist *10*, 260-268.

McDonald, A. J. (1998). Cortical pathways to the mammalian amygdala. Prog Neurobiol *55*, 257-332.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature *264*, 746-748.

Mehler, W. R., Feferman, M. E., and Nauta, W. J. (1960). Ascending axon degeneration following anterolateral cordotomy. An experimental study in the monkey. Brain *83*, 718-750.

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. Brain Res Cogn Brain Res *14*, 31-40.

Meredith, M. A., Nemitz, J. W., and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci *7*, 3215-3229.

Meredith, M. A., and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. Science *221*, 389-391.

Meredith, M. A., and Stein, B. E. (1985). Descending efferents from the superior colliculus relay integrated multisensory information. Science *227*, 657-659.

Meredith, M. A., and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol *56*, 640-662.

Merzenich, M. M., and Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res *50*, 275-296.

Mesulam, M. M., and Mufson, E. J. (1982). Insula of the old world monkey. III: Efferent cortical output and comments on function. J Comp Neurol *212*, 38-52.

Metzger, W. (1931). Versuch einer gemeinsamen Theorie der Phaenomene Froehlichs und Haszelhoffs und Kritik ihrer Verfahren zur Messung der Empfindungszeit. Psychol Forsch *16*, 176-200.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. Cognit Psychol *14*, 247-279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. Percept Psychophys *40*, 331-343.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. J Exp Psychol Hum Percept Perform *17*, 160-169.

Miniussi, C., Girelli, M., and Marzi, C. A. (1998). Neural site of the redundant target effect electrophysiological evidence. J Cogn Neurosci *10*, 216-230.

Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. Brain Res Cogn Brain Res *17*, 154-163.

Morel, A., Garraghty, P. E., and Kaas, J. H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J Comp Neurol *335*, 437-459.

Morel, A., and Kaas, J. H. (1992). Subdivisions and connections of auditory cortex in owl monkeys. J Comp Neurol *318*, 27-63.

Morris, J. S., Ohman, A., and Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. Nature *393*, 467-470.

Morris, J. S., Ohman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. Proc Natl Acad Sci U S A *96*, 1680-1685.

Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., and Burr, D. C. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. Nat Neurosci *3*, 1322-1328.

Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., and Husain, M. (2003). The anatomy of visual neglect. Brain *126*, 1986-1997. Epub 2003 Jun 1923.

Müller, J. (1826). Zur vergleichenden Physiologie des Gesichtssinnes des Menschen und der Tiere. (Leipzig, Knobloch).

Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (1996). Temporal constraints on the McGurk effect. Percept Psychophys *58*, 351-362.

Munk, H. (1881). Uber die Funktionen der Grosshirnrinde. Gesammelte Mitteilungen aus den Jahren 1877-1880. (Berlin, Hirshwald).

Murray, E. A., and Mishkin, M. (1985). Amygdalectomy impairs crossmodal association in monkeys. Science *228*, 604-606.

Myers, A. K., Cotton, B., and Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. Percept Psychophys *30*, 33-38.

Nakamura, K., Kawashima, R., Nagumo, S., Ito, K., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Kubota, K., Fukuda, H., and Kojima, S. (1998). Neuroanatomical correlates of the assessment of facial attractiveness. Neuroreport *9*, 753-757.

Nickerson, R. (1973). Intersensory facilitation of reaction time. Energy summation or preparation enhancement? Psychol Rev *80*.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. Trends Cogn Sci *6*, 387.

Nunn, J. A., Gregory, L. J., Brammer, M., Williams, S. C., Parslow, D. M., Morgan, M. J., Morris, R. G., Bullmore, E. T., Baron-Cohen, S., and Gray, J. A. (2002). Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. Nat Neurosci *5*, 371-375.

O'Craven, K. M., Downing, P. E., and Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. Nature *401*, 584-587.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci *4*, 95-102.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., and Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia *41*, 147-155.

O'Leary, A., and Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. Percept Psychophys *35*, 565-569.

Olson, I. R., Gatenby, J. C., and Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res *14*, 129-138.

Orban, G. A., Sunaert, S., Todd, J. T., Van Hecke, P., and Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. Neuron *24*, 929-940.

Pandya, D. N., and Kuypers, H. G. (1969). Cortico-cortical connections in the rhesus monkey. Brain Res *13*, 13-36.

Paradis, A. L., Cornilleau-Peres, V., Droulez, J., Van De Moortele, P. F., Lobel, E., Berthoz, A., Le Bihan, D., and Poline, J. B. (2000). Visual perception of motion and 3-D structure from motion: an fMRI study. Cereb Cortex *10*, 772-783.

Pasupathy, A., and Connor, C. E. (1999). Responses to contour features in macaque area V4. J Neurophysiol *82*, 2490-2502.

Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998). Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang *61*, 123-144.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. Neuron *36*, 767-776.

Paulesu, E., Harrison, J., Baron-Cohen, S., Watson, J. D., Goldstein, L., Heather, J., Frackowiak, R. S., and Frith, C. D. (1995). The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. Brain *118*, 661-676.

Pavani, F., Husain, M., Ladavas, E., and Driver, J. (2004). Auditory deficits in visuospatial neglect patients. Cortex *40*, 347-365.

Pavani, F., Ladavas, E., and Driver, J. (2003). Auditory and multisensory aspects of visuospatial neglect. Trends Cogn Sci *7*, 407-414.

Pavani, F., Meneghello, F., and Ladavas, E. (2001). Deficit of auditory space perception in patients with visuospatial neglect. Neuropsychologia *39*, 1401-1409.

Penfield, W., and Perot, P. (1963). The Brain's Record of Auditory and Visual Experience. A Final Summary and Discussion. Brain *86*, 595-696.

Penny, W. D. and Holmes, A. J. (2004). Random-Effects Analysis. In Human Brain Function, R. Frackowiak, ed. (Amsterdam, Elsevier), pp. 843-850.

Peretz, I., Gagnon, L., and Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition *68*, 111-141.

Petrides, M., Iversen, S. D., and Sahgal, A. (1978). The effect of selective anterior and posterior association cortex lesions in the monkey on performance of a visual-auditory compound discrimination test. Cross-modal matching and the primate frontal cortex, Cross-modal matching in the monkey after discrete temporal lobe lesions. Neuropsychologia *16*, 527-537.

Petrides, M., and Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Eur J Neurosci *16*, 291-310.

Phan, K. L., Wager, T., Taylor, S. F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. Neuroimage *16*, 331-348.

Piaget, J. (1953). The origin of intelligence in the child (London, Routledge & Keegan Paul).

Pinek, B., Duhamel, J. R., Cave, C., and Brouchon, M. (1989). Audio-spatial deficits in humans: differential effects associated with left versus right hemisphere parietal damage. Cortex *25*, 175-186.

Poezl, O. and Redlich, E. (1911). Demonstration eines Falles von bilateraler Affektion beider Occipitallappen. Wiener Klinische Wochenschrift *24*, 517-518.

Poirier, L. J., and Bertrand, C. (1955). Experimental and anatomical investigation of the lateral spino-thalamic and spino-tectal tracts. J Comp Neurol *102*, 745-757.

Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., and Mishkin, M. (2003). Functional mapping of the primate auditory system. Science *299*, 568-572.

Press, W. A., Brewer, A. A., Dougherty, R. F., Wade, A. R., and Wandell, B. A. (2001). Visual areas and spatial summation in human visual cortex. Vision Res *41*, 1321-1332.

Price, C., Wise, R., Ramsay, S., Friston, K., Howard, D., Patterson, K., and Frackowiak, R. (1992). Regional response differences within the human auditory cortex when listening to words. Neurosci Lett *146*, 179-182.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. Trans N Y Acad Sci *24*, 574-590.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. Curr Psychol Cogn *13*, 3-51.

Rademacher, J., Caviness, V. S., Jr., Steinmetz, H., and Galaburda, A. M. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb Cortex *3*, 313-329.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci *2*, 79-87.

Rasmussen, K., Heym, J., and Jacobs, B. L. (1984). Activity of serotonin-containing neurons in nucleus centralis superior of freely moving cats. Exp Neurol *83*, 302-317.

Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. Acta Otolaryngol Suppl *532*, 34-38.

Rauschecker, J. P., and Harris, L. R. (1989). Auditory and visual neurons in the cat's superior colliculus selective for the direction of apparent motion stimuli. Brain Res *490*, 56-63.

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. Science *268*, 111-114.

Rauschecker, J. P., Tian, B., Pons, T., and Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. J Comp Neurol *382*, 89-103.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform *18*, 849-860.

Reale, R. A., and Imig, T. J. (1980). Tonotopic organization in auditory cortex of the cat. J Comp Neurol *192*, 265-291.

Riddoch, G. (1917). Dissociation of visual perception due to occipital injuries, with especial reference to appreciation of movement. Brain *40*, 15-57.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res *3*, 131-141.

Robertson, I. H. (2001). Do we need the "lateral" in unilateral neglect? Spatially nonselective attention deficits in unilateral neglect and their implications for rehabilitation. Neuroimage *14*, S85-90.

Rolls, E. T., and Baylis, L. L. (1994). Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. J Neurosci *14*, 5437-5452.

Romanski, L. M., and LeDoux, J. E. (1992). Equipotentiality of thalamo-amygdala and thalamo-cortico-amygdala circuits in auditory fear conditioning. J Neurosci *12*, 4501-4509.

Romm, S. (1989). The changing face of beauty. Aesthetic Plast Surg *13*, 91-98.

Ruckmick, C. A. (1927). The rhythmical experience from the systematic point of view. Am J Psychol *39*, 355-366.

Ruff, R. M., Hersh, N. A., and Pribram, K. H. (1981). Auditory spatial deficits in the personal and extrapersonal frames of reference due to cortical lesions. Neuropsychologia *19*, 435-443.

Ryan, T. A. (1940). Interrelations of the sensory systems in perception. Psychological Bulletin *37*, 659-698.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., and Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett *127*, 141-145.

Schnupp, J. W., Dawe, K. L., and Pollack, G. L. (2005). The detection of multisensory stimuli in an orthogonal sensory space. Exp Brain Res *162*, 181-190. Epub 2004 Dec 2015.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., and Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. Int J Psychophysiol *50*, 5-17.

Schroger, E., and Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology *35*, 755-759.

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. Nature *385*, 308.

Self, M. W., and Zeki, S. (2004). The Integration of Colour and Motion by the Human Visual Brain. Cereb Cortex *22*, 22.

Seltzer, B., and Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res *149*, 1-24.

Seltzer, B., and Pandya, D. N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. Brain Res *192*, 339-351.

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., and Yantis, S. (2004). Control of object-based attention in human cortex. Cereb Cortex *14*, 1346-1357. Epub 2004 May 1327.

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. Brain Res Cogn Brain Res *14*, 147-152.

Shapiro, K. L., Raymond, J. E., and Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. J Exp Psychol Hum Percept Perform *20*, 357-371.

Shipley, T. (1964). Auditory Flutter-Driving of Visual Flicker. Science *145*, 1328-1330.

Shipp, S., and Zeki, S. (1989a). The Organization of Connections between Areas V5 and V1 in Macaque Monkey Visual Cortex. Eur J Neurosci *1*, 309-332.

Shipp, S., and Zeki, S. (1989b). The Organization of Connections between Areas V5 and V2 in Macaque Monkey Visual Cortex. Eur J Neurosci *1*, 333-354.

Shulman, G. L., d'Avossa, G., Tansy, A. P., and Corbetta, M. (2002). Two attentional processes in the parietal lobe. Cereb Cortex *12*, 1124-1131.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? Neuron *24*, 49-65, 111-125.

Slutsky, D. A., and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. Neuroreport *12*, 7-10.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., and Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: from pleasure to aversion. Brain *124*, 1720-1733.

Smith, A. P., Henson, R. N., Dolan, R. J., and Rugg, M. D. (2004). fMRI correlates of the episodic retrieval of emotional contexts. Neuroimage *22*, 868-878.

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., and Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). J Neurosci *18*, 3816-3830.

Smith, A. T., Singh, K. D., Williams, A. L., and Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. Cereb Cortex *11*, 1182-1190.

Snow, S. R. (1999). Esthetic smile analysis of maxillary anterior tooth width: the golden percentage. J Esthet Dent *11*, 177-184.

Soroker, N., Calamaro, N., Glicksohn, J., and Myslobodsky, M. S. (1997). Auditory inattention in right-hemisphere-damaged patients with and without visual neglect. Neuropsychologia *35*, 249-256.

Spelke, E. S. (1976). Infants' intermodal perception of events. Cognit Psychol *8*, 553-560.

Spence, C., Baddeley, R., Zampini, M., James, R., and Shore, D. I. (2003). Multisensory temporal order judgments: when two locations are better than one. Percept Psychophys *65*, 318-328.

Stein, B. E. (1984). Development of the superior colliculus. Annu Rev Neurosci *7*, 95-125.

Stein, B. E. (2004). Crossmodal spatial interaction in subcortical and cortical circuits. In Crossmodal Space and Crossmodal Attention, C. a. D. Spence, J., ed. (Oxford, Oxford University Press).

Stein, B. E., and Arigbede, M. O. (1972). Unimodal and multimodal response properties of neurons in the cat's superior colliculus. Exp Neurol *36*, 179-196.

Stein, B. E., Meredith, M. A., and Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. Prog Brain Res *95*, 79-90.

Stein, B. E. and Meredith, M.A. (1993). The Merging of the Senses (Cambridge, MIT Press).

Stein, E. A., Pankiewicz, J., Harsch, H. H., Cho, J. K., Fuller, S. A., Hoffmann, R. G., Hawkins, M., Rao, S. M., Bandettini, P. A., and Bloom, A. S. (1998). Nicotine-induced limbic cortical activation in the human brain: a functional MRI study. Am J Psychiatry *155*, 1009-1015.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., and Porter, N. R. (2001). When is now? Perception of simultaneity. Proc R Soc Lond B Biol Sci *268*, 31-38.

Strange, B. A., Henson, R. N., Friston, K. J., and Dolan, R. J. (2000). Brain mechanisms for detecting perceptual, semantic, and emotional deviance. Neuroimage *12*, 425-433.

Strecker, R. E., and Jacobs, B. L. (1985). Substantia nigra dopaminergic unit activity in behaving cats: effect of arousal on spontaneous discharge and sensory evoked activity. Brain Res *361*, 339-350.

Sugita, Y., and Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. Nature *421*, 911.

Sumby, W. H. and Pollack, I. (1954). J Acoust Soc Am *26*.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica *36*, 314-331.

Sunaert, S., Van Hecke, P., Marchal, G., and Orban, G. A. (1999). Motion-responsive regions of the human brain. Exp Brain Res *127*, 355-370.

Talairach, J. and Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain (New York, Thieme).

Tanaka, H., Hachisuka, K., and Ogata, H. (1999). Sound lateralisation in patients with left or right cerebral hemispheric lesions: relation with unilateral visuospatial neglect. J Neurol Neurosurg Psychiatry *67*, 481-486.

Tanaka, K. (1996). Representation of Visual Features of Objects in the Inferotemporal Cortex. Neural Netw *9*, 1459-1475.

Tawil, R. N., Saade, N. E., Bitar, M., and Jabbur, S. J. (1983). Polysensory interactions on single neurons of cat inferior colliculus. Brain Res *269*, 149-152.

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., and Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event- related potential (ERP) recordings. Brain Res Cogn Brain Res *14*, 106-114.

Thornhill, R., and Gangestad, S. W. (1999). Facial attractiveness. Trends Cogn Sci *3*, 452-460.

Tremblay, L., and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. Nature *398*, 704-708.

Turner, B. H., Mishkin, M., and Knapp, M. (1980). Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. J Comp Neurol *191*, 515-543.

Ulrich, R. and Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. Perception and psychophysics *39*, 248-259.

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In Analysis of visual behaviour, D. J. Ingle, Goodale, M. A. and Mansfield, R. J. W., ed. (Cambridge, MIT).

Vaadia, E., Benson, D. A., Hienz, R. D., and Goldstein, M. H., Jr. (1986). Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. J Neurophysiol *56*, 934-952.

Vallar, G. (2001). Extrapersonal visual unilateral spatial neglect and its neuroanatomy. Neuroimage *14*, S52-58.

Vallar, G., Guariglia, C., Nico, D., and Bisiach, E. (1995). Spatial hemineglect in back space. Brain *118*, 467-472.

Vallar, G., and Perani, D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. Neuropsychologia *24*, 609-622.

van Beers, R. J., Wolpert, D. M., and Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. Curr Biol *12*, 834-837.

Vandenberghe, R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2001). Location- or feature-based targeting of peripheral attention. Neuroimage *14*, 37-47.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., and Orban, G. A. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science *298*, 413-415.

Verrey, L. (1888). Hémiachromatopsie droite absolue. Arch Ophtalmol (Paris) *8*, 289-301.

Vogels, R. (1999). Effect of image scrambling on inferior temporal cortical responses. Neuroreport *10*, 1811-1816.

von Economo, C. H., L. (1930). Ueber Windungsrelief, Masse und Rindenarchitecktonik der Supertemporalflaeche, ihre individuellen und ihre Seitenunterschiede. Z Neurol Psychiatr *130*.

Walker, A. E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. J Comp Neurol *73*, 59-86.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., and Schirillo, J. A. (2004). Unifying multisensory signals across time and space. Exp Brain Res *158*, 252-258. Epub 2004 Apr 2027.

Wallace, M. T., and Stein, B. E. (2000). Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. J Neurophysiol *83*, 3578-3582.

Watanabe, J., and Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. Brain Res Bull *26*, 583-592.

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., and Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. Cereb Cortex *3*, 79-94.

Welch, R. B., DuttonHurt, L. D., and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. Percept Psychophys *39*, 294-300.

Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. Psychol Bull *88*, 638-667.

Werner, H. (1934). L'Unité des Sens. Journal de Psychologie *31*, 190-205.

Whitfield, I. C., and Evans, E. F. (1965). Responses of Auditory Cortical Neurons to Stimuli of Changing Frequency. J Neurophysiol *28*, 655-672.

Whitney, D. (2002). The influence of visual motion on perceived position. Trends Cogn Sci *6*, 211-216.

Wickelgren, B. G. (1971). Superior colliculus: some receptive field properties of bimodally responsive cells. Science *173*, 69-72.

Williams, T. M., and Aitken, L. S. (1977). Development of pattern classification: auditory visual equivalence in the use of prototypes. Developmental Psychology *13*, 198-204.

Wojciulik, E., and Kanwisher, N. (1999). The generality of parietal involvement in visual attention. Neuron *23*, 747-764.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex *13*, 1034-1043.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., and Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. Nat Neurosci *5*, 995-1002.

Yen, C. T., and Blum, P. S. (1984). Response properties and functional organization of neurons in midline region of medullary reticular formation of cats. J Neurophysiol *52*, 961-979.

Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. Neuropsychologia *23*, 31-41.

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. Cereb Cortex *11*, 946-953.

Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci *14*, 1908-1919.

Zatorre, R. J., and Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. Neuropsychologia *31*, 221-232.

Zeki, S. (1990). A century of cerebral achromatopsia. Brain *113*, 1721-1777.

Zeki, S. (1991). Cerebral akinetopsia (visual motion blindness). A review. Brain *114*, 811-824.

Zeki, S. (1993). A vision of the brain (Oxford, Blackwell Scientific Publications).

Zeki, S. (1999). Inner Vision (Oxford, Oxford University Press).

Zeki, S., and Bartels, A. (1999). Toward a theory of visual consciousness. Conscious Cogn *8*, 225-259.

Zeki, S., Perry, R. J., and Bartels, A. (2003). The processing of kinetic contours in the brain. Cereb Cortex *13*, 189-202.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. J Neurosci *11*, 641-649.

Zeki, S. M. (1969). Representation of central visual fields in prestriate cortex of monkey. Brain Res *14*, 271-291.

Zeki, S. M. (1971). Convergent input from the striate cortex (area 17) to the cortex of the superior temporal sulcus in the rhesus monkey. Brain Res *28*, 338-340.

Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. Brain Res *53*, 422-427.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol *236*, 549-573.

Zeki, S. M. (1977). Colour coding in the superior temporal sulcus of rhesus monkey visual cortex. Proc R Soc Lond B Biol Sci *197*, 195-223.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. Nature *274*, 423-428.

Zihl, J., von Cramon, D., and Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. Brain *106*, 313-340.

Zihl, J., von Cramon, D., Mai, N., and Schmid, C. (1991). Disturbance of movement vision after bilateral posterior brain damage. Further evidence and follow up observations. Brain *114*, 2235-2252.

Zimmer, U., Lewald, J., Erb, M., Grodd, W., and Karnath, H. O. (2004). Is there a role of visual cortex in spatial hearing? Eur J Neurosci *20*, 3148-3156.